

Summits that matter

The European Commission has made good progress in gathering support for its new programme of basic and applied research. Now Europe's industries and heads of state need to fulfil promises made two years ago.

Few occasions would seem to be more remote from the everyday concerns of researchers than meetings of heads of state. Give researchers the prospect of significant funds to pursue their interests, and the autonomy to do it in the way they think is best, and you'll get their attention. Issue summit communiqués about the need to increase competitiveness and they'll nod off. But both approaches are now on the agenda in Europe, and both have the potential to boost scientists' scope for action.

No doubt one such summit, the Lisbon European Council in 2000, had little impact on scientists at the time. Certainly, it has made precious little difference to them since, even though it was intended to mark the start of a golden age for science and innovation in the European Union (EU). Within a decade, the heads of state asserted, the EU would become the most dynamic, competitive, sustainable, knowledge-based economy in the world, enjoying full employment and social cohesion. Massive investment in science and education would ensure the realization of this vision.

Two years later, any euphoria has vanished. Key deliverables, such as a cheap EU patent, are still yet to materialize, despite much discussion. Negotiations over the planned European satellite navigation system Galileo, a major public-industry technology investment, are bogged down. The 'European Research Area', a concept pushed by research commissioner Philippe Busquin, suffers from a lack of support from some member states and — perhaps as a consequence — a lack of substance. Insiders at the European Commission openly speak of a "delivery gap" which threatens its credibility. And industry's expenditure on research and development itself is too low: only 55% of R&D in the EU is funded by industry, compared with 72% in Japan and 67% in the United States.

Overcoming obstacles

But things are looking up. Supported by the zealous Spanish presidency of the EU, the commission is now working hard to get back on track. Ministers for industry and for research from 15 member states will meet next week in Spain for the first time, to discuss obstacles to the Lisbon objectives. Breakthroughs are unlikely, but this is nonetheless a long-overdue initiative.

The commission's 2003–06 Framework programme for research, now well on its way to final approval to the tune of 16.2 billion euros (US\$14 billion), will strongly support industrial participation in large 'integrated' projects (see <http://europa.eu.int/comm/research/nfp.html>). Such projects are yet to be defined by the research community and look daunting to manage, but are of sufficient scale and scope to make an impact on both fundamental and applied research, and are already attracting interest. And at the European Council in Barcelona in March, the commission will urge the heads of the member states to set a target of 3% of gross domestic product for the overall level of public and private spending on research by 2010, two-thirds of which would be contributed by industry — double their present level.

Such a goal cannot be achieved by government or commission fiat. Nor can it be expected to happen spontaneously: Finland and

Sweden, which already surpass the 3% target, hardly serve as a model for large and highly diversified economies such as those of Germany, Italy, France or Britain. Thus, one must move step-by-step to create an economic and regulatory environment that is more friendly to companies of all sizes willing to invest in research, and to encourage some multinational companies to shift the balance of their R&D spending towards labs in Europe.

Action plan

The European Commission estimates that the market potential of the biotechnology industry in Europe will be 100 billion euros by 2005. But the exploitation of this potential is sluggish. Accordingly, at the Barcelona summit, the commission will present an action plan for the life sciences, which it published last week, including the development of a bioinformatics infrastructure, regulatory harmonization, and centralized authorization procedures for the environmental release of genetically engineered organisms (GMOs).

The plan also reflects the commission's need to address the tensions between the economic benefits of biotechnology and the concerns of Europe's citizens, in particular with regard to GMOs, whose commercial release in Europe is currently under a moratorium. It calls for broad public debate, encourages regular stakeholder forums and promises the establishment of a publicly accessible 'molecular register' of occurrences of deliberate genetic modification. With regards to ethical controversies in biomedical research, it suggests networking of national and local ethics bodies and encourages self-regulatory initiatives in the scientific community and industry. At the same time, it is pursuing new pro-pharmaceutical legislation, such as accelerated approval procedures and one-year conditional authorization.

Such regulatory changes will require determined political leadership. From a scientist's point of view, any attempt to increase research spending and temper regulatory burdens is to be welcomed. So, too, is another key aspect of the new Framework programme: large-scale networks, each funded with as much as 20 million euros over five or more years, with money that is not earmarked and therefore provides autonomy for the collaborators in their ability to develop new facilities or new centres of excellence.

Underlying these plans is a determination by Philippe Busquin to move away from non-scientific agendas and to foster world-beating collaborations. Gone from his plans is "cohesion", whereby laboratories from less-well-developed countries are included in collaborations thanks to affirmative action. There is also a stronger move than hitherto to support the pursuit of scientific knowledge for its own sake, as well as for its application.

That shift, combined with giving researchers more independence and, as one commission insider puts it, an ambition to distribute funds like hailstones rather than fine mist, should all be good news to outstanding researchers able to make use of the opportunities. But if the heads of state don't also provide high-level backing, both for the commission and within their own countries, Europe's technological and, in the longer term, scientific competitiveness seem certain to weaken. ■





Ship shape
Wreck preservation project makes a splash
p460



Parting of ways
Celera affirms focus on drug discovery in wake of Venter's exit
p461



Health check
Side-effects bring trial for Alzheimer's drug to a standstill
p462



Up and away
NASA balloon undergoes record-breaking flight
p465

Bush's budget boost puts NIH on target for doubled figures

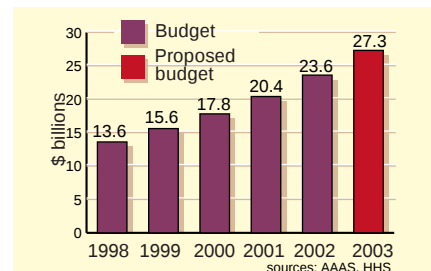
Erika Check, Washington

The National Institutes of Health (NIH) is set to complete a five-year spurt of growth next year, with its funding rising to an eye-popping \$27.3 billion in the fiscal year 2003.

US health secretary Tommy Thompson announced on 26 January that President George W. Bush will request a \$3.7-billion increase for the biomedical research agency in his 4 February budget proposal. This would fulfil a drive initiated in Congress — and endorsed by Bush in his election campaign — to double in five years the NIH budget from its 1998 level of \$13.6 billion.

The drive has been led by patient advocacy groups, science lobby groups including Research!America, and legislators such as John Porter, the former chair of the House appropriations subcommittee in charge of health spending, who retired from Congress in 2000. The success of the effort has been watched enviously by scientists in other disciplines and in other countries.

Congress is widely expected to approve the president's request when it completes the 2003 budget this summer. Bush is asking for \$5.5 billion next year for cancer research, up from \$4.9 billion this year; and \$1.5 billion



On the rise: the past five years have seen funding flood into the National Institutes of Health.

for research related to bioterrorism — five times this year's expenditure. Most of the new bioterrorism money will go to the National Institute of Allergy and Infectious Diseases (NIAID). Rumours that NIAID director Anthony Fauci is about to be appointed to the vacant directorship of the NIH reached a new intensity ahead of Bush's State of the Union address to Congress on 29 January.

Fauci would not comment on the rumours, but says that his institute would spend the bioterrorism money on basic research, drug and vaccine discovery, clinical research and construction of new facilities.

With two-fifths of the new money going to the NIAID, Bush's proposal will see the other 26 NIH institutes enjoying somewhat smaller increases than in recent years. But Fauci pledges that the expansion of bioterrorism funds would also boost research on diseases such as cancer, malaria, tuberculosis and AIDS.

Jordan Cohen, president of the Association of American Medical Colleges, welcomes the funding announcement. Looking ahead, he says: "It's probably not reasonable to expect funding to grow at exactly the same rate it has, but meeting the scientific opportunities that exist is going to require continued investment."

April Burke of Lewis-Burke Associates, a lobbying group that works with universities and research organizations, says: "There's a lot of money going to cancer and bioterrorism, but the important thing is that the administration is supporting the overall increase of the NIH, and that's excellent."

Physicists head for collision course

Geoff Brumfiel, Washington

The construction of a linear electron-positron collider should be the top priority for high-energy physics over the next 20 years, an advisory panel has told the US government.

The panel, a sub-group of the High Energy Physics Advisory Panel that is

chaired by Jon Bagger of Johns Hopkins University, also recommends that the United States should seek to host the collider.

But if it is built, the machine will cost between \$5 billion and \$7 billion, and will require strong collaboration between physicists in Japan, the United States and Europe. It remains unclear what form that collaboration will take.

So far, regional groups of physicists have done most of the conceptual design work on the collider, and Bagger says that an international organization is needed "almost immediately" to improve the way these efforts are coordinated.

"The international community has to get its act together," agrees Albrecht Wagner, director of DESY, Germany's high-energy physics research centre in Hamburg. He adds that any collaboration will face tough choices about where to build the machine and which technologies to use.

US high-energy physicists acknowledge that they face an uphill struggle in convincing the Department of Energy and the National Science Foundation, the agencies that support their discipline, to back the project. But Bagger is hopeful. "A 10–30% increase in our budget would do the job," he says. "That's not unreasonable to contemplate."

A klystron such as this would accelerate high-energy microwaves for the planned linear collider.



Minister set to reform Australia's universities

Peter Pockley, Sydney

Australia's hard-pressed universities can expect no financial relief from the new government until May of next year at the earliest, says Brendan Nelson, the country's new minister for education, science and training.

But Nelson, a physician and former president of the Australian Medical Association who looks set to bring considerable energy and influence to his position, pledged to enact reforms that will help Australian universities to carry out internationally competitive research.

"The situation is not sustainable without reform," says Nelson. He says that he will immediately set up a consultative panel to help him prepare a university-reform package, for consideration by Australia's cabinet within a year. The reforms may open up the possibility of extra funding in the May 2003 budget, he says.

"The universities have been reviewed to death," says Nelson. "We know what the problems and challenges are." The new panel will avoid revisiting the ground covered in previous reviews, he says, and will instead propose concrete changes in university governance, working conditions and the way in which specialist strengths are split between universities. Nelson wants to see "one or two universities become world-class", but

acknowledges that this is "not likely under current funding arrangements".

After its November re-election, the Liberal Party government of Prime Minister John Howard shifted responsibility for Australia's science agencies from the department of industry to Nelson's department. Nelson pledges to appoint a scientific adviser in his office "who will have credibility in the science community and will live and breathe science".

Nelson has wasted little time in challeng-



Facing the challenge: Brendan Nelson says he wants to concentrate on concrete changes.

ing the universities to address their own shortcomings: at a recent meeting with vice-chancellors, for example, he told them that they had lost the support of "the everyday person". Last year, the vice-chancellors lobbied for an extra A\$1 billion (US\$515 million) a year to make up for cuts made since Howard came to power in 1996, but Nelson has made it clear that they will have to settle for less. The government is committed to an "innovation package" containing A\$2.9 billion in new funding over five years (see *Nature* **409**, 655; 2001), but has delivered only A\$160 million for this year.

Nelson was once something of a maverick in Howard's staunchly conservative government — he wore an earring when he first entered parliament in 1996. Now, however, as the measured and articulate face of the government's science policy, his style contrasts with the abrasiveness of his predecessor, David Kemp, whose own reform package for the universities was rapidly disowned by the government (see *Nature* **402**, 113; 1999). Nelson's openness to discussion has been welcomed by scientific societies. The Australian Academy of Science and the Federation of Australian Scientific and Technological Societies have already drawn up their own lists of ideas for the new minister and his consultative group to chew on. ■

P. POCKLEY

Undersea plan leaves wrecks to rest in peace

Quirin Schiermeier, Munich

Archaeologists in northern Europe are planning a fresh approach to the study and conservation of shipwrecks.

Under a project backed by the European Union, researchers will investigate the wrecks *in situ* and use technology to let the public explore virtual images of them, without the expense and physical disruption of hauling the structures to the surface.

"Shipwrecks are best studied and preserved in their natural environments," says Sallamaria Tikkanen, a curator at the Maritime Museum of Finland, which will coordinate the project.

The idea of opening shipwrecks up to the public but leaving them in place gained momentum after the discovery in 1999 of the wreck of the *Vrouw Maria*. This eighteenth-century trade ship, which sank off the coast of Finland, is particularly well preserved. It will now be left in place and studied as part of the new project.

"Shipwrecks are cultural possessions, and we have a responsibility to preserve them for future generations," says Hauke Jöns, a conservationist at the Archaeological State Museum of Mecklenburg-Vorpommern in



Four ships selected for study were seen as representative of the wrecks off northern Europe's coasts.

Germany. "But they should not be removed from the place where they have met their fate."

Other shipwrecks included in the project are the *Erik Nordevall*, a nineteenth-century paddle steamer that sank in Sweden's Lake Vanern; the *Burgzand Noord* in the Wadden Sea off Holland; and the *Darsser Kogge*, a medieval ship discovered in the sea close to the German town of Prerowstrom, near Rostock.

The project includes research on the process of degradation and on the problem of *Teredo navalis*, a shipworm that feeds on

timber. The project will also explore various conservation methods, such as covering shipwrecks in sarcophagus-like tarpaulins and using sandbags to protect them from undersea currents.

Equipment at each site will measure salinity, temperature, water movement and the prevalence of microorganisms, warning the researchers of changes that could cause the wrecks to deteriorate.

The shipwrecks will be extensively filmed, and the resulting images will allow Internet users and museum visitors to take 'virtual tours' of them. ■

Venter's departure sees Celera seek therapies

Carina Dennis

The announcement on 22 January that Craig Venter was stepping down as president of Celera provided few clear pointers about the future of either the maverick geneticist or the biotechnology operation, based in Rockville, Maryland. But according to observers of both, the man may have a better chance of landing on his feet than the company he built.

Celera's revenues are projected to reach \$130 million this year, as more academic and commercial users subscribe to its genome databases. That will make the company profitable, but it won't meet investors' heady expectations. So, for the past year or so, Celera has been pursuing a new business plan based on drug discovery.

But Tony White, chief executive of Celera's parent company, Applera, and now acting president of Celera itself, acknowledges the company's lack of expertise in drug discovery. He says there will now be a "top-to-bottom review" of the company's operations, and that it will hire the skills it will need. In this endeavour, Celera's greatest asset may be its cash reserve of almost \$1 billion, accrued when investors' hopes for the company peaked two years ago.

Venter's departure leaves the door open for White to recruit a new leader for Celera with the right experience in drug discovery and development. The company has reorganized to focus on the systematic identification of proteins that are expressed differently in diseased tissue and normal tissue, and which may represent good drug targets.

But lists of poorly characterized potential drug targets have little value in themselves. In November, Celera bought Axys Pharmaceuticals, a small-molecule drug development company based in San Francisco, whose capabilities in pharmacology, assay development and structural biology will help it to select the most promising drug targets.

Samuel Broder, Celera's chief medical officer, says this results in a "unique combination of proteomics, combinatorial chemistry and computational biology" that will give Celera an edge in drug discovery. The company says it has already identified a couple of promising compounds for the possible treatment of autoimmune disease and thrombosis.

But sceptics point out that neither Celera nor Axys has an established track record in this area. "Celera is one of perhaps 500 to 1,000 raw start-up companies interested in drug discovery that is in a rudimentary and undefined state," says William Haseltine, a former colleague and long-time critic of Venter. Haseltine is chief executive of Human Genome Sciences, another Rockville-based



On the hunt: Craig Venter leaves Celera competing with many companies in the search for new drugs.

biotechnology company. He adds that any small company "that thinks it can outdo large companies in chemical-based drug discovery" is labouring under a major misconception.

Venter's departure was not entirely unexpected, and had little impact on Celera's stock. It held steady at a value of around \$23 during the week of the announcement — far below its March 2000 high-water mark of \$247.

Celera's sister company, Applied Biosystems, has suffered from declining sales of

its sequencing machines and reagents, which it attributes mainly to lower-than-expected demand from genomics researchers.

As for Venter himself, the normally volatile *bête noire* of US molecular biology was unavailable for comment, reportedly sunning himself on his racing yacht in the Caribbean.

According to Broder, Venter is "returning to his first love: science". There's no word yet on when or where Venter will resurface, but as Broder puts it: "There are always creative surprises whenever Craig is involved." ■

Senators square up over cloning

Erika Check, Washington

With the US Senate set to debate legislation on human cloning over the next few weeks, senators are lining up behind two rival positions on the issue.

On 24 January, senators Tom Harkin (Democrat, Iowa) and Arlen Specter (Republican, Pennsylvania) introduced a bill that would allow cloning of human embryos for research purposes, but ban its use in human reproduction.

This legislation is similar to a bill introduced last month by Dianne Feinstein (Democrat, California). For jurisdictional

reasons, Feinstein's bill is more likely to make it to the Senate floor, where it will compete with a proposal from Sam Brownback (Republican, Kansas) that would ban cloning for any purpose.

Most research organizations, including the Federation of American Societies for Experimental Biology and the Association of American Medical Colleges, have backed the Feinstein bill.

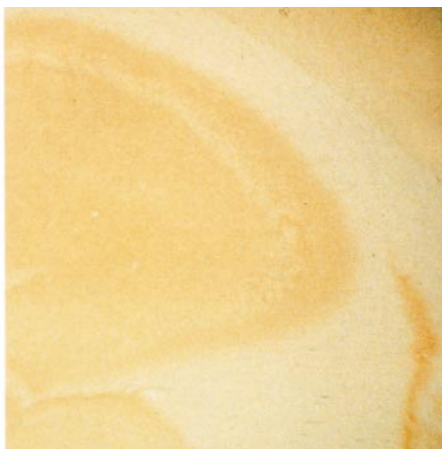
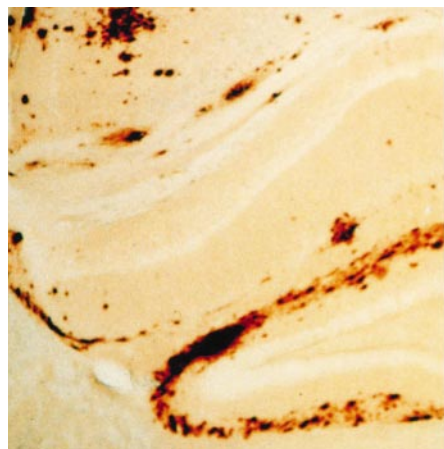
The Senate has yet to take a vote on the politically charged issue of cloning, although Senate majority leader Tom Daschle (Democrat, South Dakota) has responded to increasing pressure to do so by promising to allow time for a debate during February or March. The House of Representatives passed a ban on cloning for research or reproduction last July by a margin of 265 to 162.

But Congress may still fail to agree on legislation. The Senate is thought to be closely divided on the issue, but if it passes the Feinstein bill, it is highly unlikely that House and Senate negotiators would be able to agree a compromise on the basic issue of whether cloning should be permitted for research purposes. ■



Putting proposals: (from left to right) Tom Harkin, Dianne Feinstein and Sam Brownback.

Nerve inflammation halts trial for Alzheimer's drug



Before and after: plaques of amyloid- β (left) in mouse brain can be treated with the drug AN-1792.

Erika Check, Washington

Trials of a new immunotherapy drug for Alzheimer's disease have hit a roadblock, with four of the participants developing potentially serious problems in the nervous system. Some researchers say the side-effect could have been anticipated before the trial began.

The drug's manufacturer, Elan of Dublin in Ireland, and its collaborating company, Wyeth-Ayerst of Madison, New Jersey, announced on 18 January that they were temporarily suspending a 360-patient trial of the drug, AN-1792, being carried out in the United States and Europe, while they investigate the problem. The four patients, all based in France, show "clinical signs consistent with inflammation in the central nervous system", the statement says.

Max Gershenoff, a spokesman for Elan, declined to elaborate on what led the investigators to diagnose the inflammation. He says the company hopes that the safety-monitoring committee that has been asked to establish the source of the inflammation will do so before the patients receive their next scheduled dose of the drug. A Wyeth-Ayerst spokesman says that one of the patients has already recovered.

Three years ago, Elan reported that AN-1792 attacks tangles of human amyloid- β protein, called plaques, in the brains of transgenic mice (see D. Schenk *et al. Nature* **400**, 173–177; 1999). These plaques are diagnostic of Alzheimer's disease. Elan began phase I trials (for human safety) in December 1999 and reported no ill effects in 80 recipients of the drug. Last October, the company began its phase IIA trial to evaluate the drug's effectiveness against the symptoms of Alzheimer's.

But some scientists warn that AN-1792 could cause a dangerous autoimmune response in humans. The drug contains a

synthetic version of amyloid- β , and causes the body's immune system to attack amyloid- β plaques. But scientists say the autoimmune response might also target amyloid precursor protein, which gives rise to amyloid- β . If it does, the side-effects could be devastating, because the precursor protein is found in many normal cells, including neurons.

"Some people have predicted that inflammation would be a side-effect," says Stephen Robinson, a neuroscientist at Monash University in Victoria, Australia.

Robinson says he is concerned that Elan has not reported a crucial safety test of its vaccine. The company tested the vaccine in transgenic mice that had brain plaques made of human amyloid- β proteins. But mice have their own form of amyloid- β and amyloid precursor protein. Elan has not published information on whether innoculating mice with mouse amyloid- β causes them to mount an immune response against their own amyloid precursor protein, says Robinson.

"What you're doing in the transgenic mice is simply clearing human amyloid from their brains and leaving the mouse amyloid intact," Robinson says. "The key experiment would be to take wild-type mice and inoculate them against their own mouse amyloid- β and amyloid precursor protein."

Gershenoff would not comment on whether those tests had been carried out. "We haven't made any announcement about that," he says.

Researchers say that, as Elan has already tested AN-1792 in 360 patients in clinical trials in the United States and Europe, it may be reluctant to blame the drug for the inflammation in the French patients. "It's premature to be overly concerned about this, given the small number of patients involved," says Karen Duff, an associate professor of psychiatry at New York University. ■

Progress in human genetics hindered by reluctance to share

David Adam

Data sharing in human-genetics research is being hampered by concerns over intellectual property rights, says one of the first studies to try to quantify the problem.

Of the 1,240 human geneticists surveyed by a team led by Eric Campbell of the Institute for Health Policy at Massachusetts General Hospital, Boston, almost half complained that a co-worker had refused them access to data, information or materials in the past three years.

According to 28% of the survey respondents, such refusals meant that they could not replicate published work, and 58% felt that their own research had suffered as a result. The survey's findings were published on 23 January in the *Journal of the American Medical Association* (287, 473–480; 2002).

Requests for materials, such as cell lines or the 'knockout' mice used to help to identify gene function, were refused more often (35%) than requests for extra data (25%) or methods (16%).

"It may be that material-transfer agreements have become so complex and demanding that they inhibit sharing," says Campbell. Universities often require researchers to complete copious paperwork to ensure that the institution retains the intellectual property rights to any work deriving from the materials before they are handed over.

Joann Boughman, executive vice-president of the American Society of Human Genetics, agrees. "Transfer agreements are now horrendously complicated," she says. "They used to be simple exchanges between scientists, but they have now evolved into complex business arrangements."

Four-fifths of those who admitted to not honouring a request cited the time and effort needed to locate or produce materials as a reason.

"The difficulty is time," agrees Kay Davies, a human-genetics researcher at the University of Oxford. "In the early days of our work we were getting three or four requests a day for probes and producing them is almost a full-time job."

Despite the scale of the problem, researchers contacted by *Nature* denied that there is a culture of secrecy in human genetics. "If anything is delaying progress in the field, it is changing attitudes towards intellectual property, rather than a reluctance to share," says Robert Ferrell, chair of human genetics at the University of Pittsburgh in Pennsylvania. ■

Survey overlap impedes fossil hunters' study

Rex Dalton, San Diego

A dispute is simmering between American- and Italian-led teams over access to one of the world's most promising hominid fossil fields.

The two groups are each feeling their way through research regulations in the East African state of Eritrea, which gained its independence from Ethiopia in 1993 and has since been attempting to open up its rich cultural heritage to scientific investigation.

The dispute is over the northern Danakil Depression, where rocky outcrops have brought fossils to the surface to reveal tantalizing clues of early hominids in the region (see E. Abbate *et al. Nature* **393**, 458–460; 1998).

Since 1995, a team from the University of Florence, Italy, has been exploring a region around the hamlet of Buia, where a million-year-old hominid skull was found. But when the Italian team, led by Lorenzo Rook, returned earlier this month to begin new explorations there, they were told that their research permits would be delayed.

According to Rook, Eritrean officials blamed the delay on problems that had arisen last October during a field trip in the same region by a team led by Randall Susman of Stony Brook University in New York. The Italian researchers are also annoyed that Susman was covering territory where they had been previously working.



Disputed territory: the Danakil Depression, where two teams' explorations overlapped.

Such overlaps are frowned on by most physical anthropologists, but they are permitted by Eritrean regulations, which prohibit "exclusive" territorial concessions.

The field trip by Susman, a Stony Brook anatomist with 25 years' experience in functional-morphology research in Africa, has sparked unease in the Eritrean capital of Asmara. It has also caused ripples in San Francisco, home of the Leakey Foundation, which supports research into human origins and has provided US\$15,000 to pay for the expedition.

In Asmara, Eritrean government officials have suspended Susman's survey permit, after his October field trip had included a South African geologist not approved as a member of the team. Susman acknowledges that this was an "administrative mistake", saying it was an unintentional violation of regulations.

In San Francisco, Leakey Foundation officials criticized Susman for entering the Italian team's territory, when his grant award specifically stated that "the Leakey Foundation wishes to emphasize that your and the Italian team's study area should not overlap".

Clark Howell, chairman of the Leakey Foundation's scientific executive committee and a former professor of anthropology at the University of California, Berkeley, has suggested that Susman should return the grant to the foundation in light of the incident — but the full executive committee has declined to ask for such a refund.

Susman says that he tried hard to avoid the Italians' region, but ended up there because of inadequate maps, difficulties in securing Eritrean scientific assistance and confusion over the regulations. "I made every reasonable effort to prevent this from happening," he says, "but it was impossible to avoid."

Fossils collected on the expedition have been deposited at the Eritrean National Museum, Susman says. But Tewelde Medin Teclé, an Eritrean geologist who has worked with the Italian researchers, has accused Susman's team of harming the site by removing fossils and ancient tools. Susman denies this.

Late last week, Rook said that he hoped to leave soon for the field near Buia. "We are paying for the mistake someone else made," he says.

Delegates nudge fusion project closer to reality

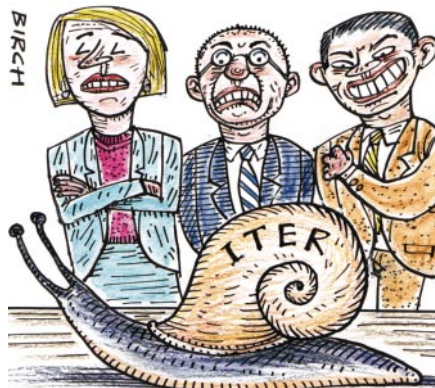
David Cyranoski, Tokyo

International negotiators met in Tokyo last week to discuss plans for the construction of ITER, the proposed magnetic-fusion energy experiment.

Among the topics discussed by delegates from the European Union, Russia, Canada and Japan at the two-day negotiations were procedures for selecting a site for the experiment and for procuring equipment.

The United States, which is considering rejoining the project (see *Nature* **415**, 247; 2002), did not take part in the negotiations. But Robert Goldston, director of the Princeton Plasma Physics Laboratory, the largest magnetic-fusion laboratory in the United States, did attend an associated symposium that was held to bolster public support for the project. "I hope the United States joins," he says.

Whether Japan itself will offer to host the experiment remains unclear. In a



SNAIL RACING IN JAPAN.

preliminary report issued last December, the Council for Science and Technology Policy, Japan's highest scientific research body, reiterated its desire to host the project. The report said that this would cost Japan ¥700 billion (US\$5 billion) over the first

10 years, for construction and operations. Japan's estimated cost of participation if the experiment is built elsewhere is ¥300 billion.

Within Japan, public support for nuclear energy has recently weakened (see *Nature* **411**, 729; 2001), and anti-nuclear groups used the negotiations as an opportunity to raise concerns about ITER's safety and viability.

Japan's Ministry of Education, Science, Sports and Culture is still assessing two proposed sites for the facility. One of them, at Naka in Ibaraki, would be close to the Japan Atomic Energy Research Institute's existing nuclear-fusion facilities and quite close to Tokyo's main airport. The second site, at Rokkasho in Aomori, is on the northern tip of Japan's main island, where much of the country's nuclear waste is stored.

The ministry hopes to choose between the two sites before the next ITER meeting in Moscow on 19 March.

New rules for German junior professors 'put jobs under threat'

Munich Almost 6,000 young scientists in Germany have signed a petition against a proposed new university law, which researchers fear could lead to the loss of thousands of temporary academic jobs.

The new law is being introduced to replace the existing *Habilitation* qualification, which requires scientists to complete a second research thesis after obtaining a PhD. Science minister Edelgard Bulmahn plans instead to introduce a system of six-year junior professorships, and to limit the time that scientists are allowed to work on fixed-term contracts to 12 years (see *Nature* **415**, 257–258; 2002). In exceptional cases, universities will still be allowed to issue limited contracts after this period. But Thomas Mergel, spokesperson for the young researchers, says that universities are unlikely to do this because scientists could then sue them for a permanent job.

The law is due to come into force in the next few weeks, although a parliamentary science committee of the ruling Social

Democratic Party says that it recognizes the problem and is calling for an investigation.

♦ www.wissenschaftlichernachwuchs.de

Weapons facilities riddled with risk, says congressman

Washington Safety and security flaws at US nuclear-weapons facilities are highlighted in a letter written by US congressman Edward



Edward Markey: public misgivings.

Markey (Democrat, Massachusetts) to energy secretary Spencer Abraham.

Markey claims that government officials posing as terrorists were able to gain access to weapons-grade plutonium or uranium in over half of the exercises conducted to assess security at the facilities. He cites a 1997 exercise in which mock terrorists were able to steal weapons-grade uranium using a garden cart. Markey also states that storage vaults for the radioactive materials are would not protect against aircraft impact or large bombs, and that some were left unlocked.

John Gordon, head of the National

Nuclear Security Administration, the agency in the energy department that runs the weapons programme, claims that Markey's interpretation is misleading. "Our forces... are tested by outside challengers, often to failure, so we know where our weaknesses are," he says. "Then we fix the problem."

Computer 'trespasser' escapes jail sentence

Atlanta A computer administrator charged with computer theft and computer trespass for downloading distributed software onto computers at his workplace will not serve the jail sentence with which he had been threatened.

David McOwen's former employers at DeKalb Technical College in Atlanta, Georgia, accused him of downloading software for tackling mathematical problems such as encryption without their permission (see *Nature* **415**, 107; 2002). The case has now been settled out of court. McOwen will pay around US\$2,000 for the cost of removing the software from the machine, and will complete a term of community service. Georgia state prosecutors had originally sued for US\$415,000 to compensate for bandwidth use.

Distributed software, such as SETI@home, which aids the search for

extraterrestrial life, divide time-consuming computer problems between many different machines. McOwen was the first person known to have faced a criminal charge for installing such software.

Grant cuts loom for many UK research departments

London Many British university departments are facing cuts in the grants that they receive to support their research work, after the Higher Education Funding Council for England (HEFCE) decided how to allocate funds on the basis of its recent Research Assessment Exercise.

HEFCE uses the results of the annual exercise to distribute £940 million (US\$1.3 billion) of government funding to support university infrastructure. But with over 40% of departments obtaining the highest 5 or 5* ratings last year (see *Nature* 414, 834; 2001), HEFCE was forced to reconsider the formula that it normally uses to calculate grant sizes.

The council announced on 23 January that its funding for departments rated 4 and 5 will be cut by 30% and 14%, respectively, to reward fully those that achieve 5*. Departments that score 3b or less will receive nothing.

♦ www.hefce.ac.uk

US plan fuels nuclear proliferation fears

Washington Plans to convert some of the US stockpile of weapons-grade plutonium into fuel for nuclear reactors have angered groups that oppose nuclear proliferation.

The Department of Energy had been considering two plans to reduce the stockpile: conversion to fuel, and 'immobilization', whereby the plutonium is rendered unfit for use by encasing it in glass and storing it underground. The latter option has been abandoned, and 34 tonnes of plutonium will now be converted into mixed-oxide fuel.

Advocates of the plan claim that abandoning immobilization will save billions of dollars and fulfil treaty obligations with Russia to cut the stockpile. But opponents argue that the strategy will increase risks by releasing the plutonium into the commercial

world, and could encourage other countries to pursue similar plans.

Record-breaking Antarctic balloon is enduring prospect

Washington A scientific balloon launched by NASA has set a new endurance record after clocking up a 32-day flight over the South Pole.

The balloon, part of the trans-iron galactic element recorder (TIGER) experiment investigating cosmic rays, spent most of its journey at a height of 38 kilometres. NASA hailed the achievement as an important milestone towards the goal of 100-day balloon flights as alternatives to satellite missions.

Long-duration balloon flights could soon be extended to the Northern Hemisphere. NASA says the United States and Russia are close to agreement over using Russian airspace for such missions.

♦ cosray2.wustl.edu/tiger



Pole star: NASA's balloon spent 32 days in flight over Antarctica, setting a new endurance record.

TIGER

The rise of the bean counters

The Office of Management and Budget always exerts a powerful influence over the US government. But it is now taking a higher profile, and a stronger interest in science. Should researchers be alarmed? Colin Macilwain reports.

People who work for the White House Office of Management and Budget (OMB) don't expect to be loved. Responsible for managing the federal government's vast bureaucracy and monitoring its spending, the OMB is widely seen by those it oversees as the devil in disguise. "They're looking for the horns and the tail as soon as you walk in the door," says one former OMB official, who used to keep tabs on the National Institutes of Health.

Adored they may not be, but OMB staff are certainly powerful — especially in fram-

ing the president's annual budget request to Congress. Under Mitch Daniels, appointed as OMB director a year ago, the office has risen to even greater prominence and, some would say, notoriety. As its officials pore over their portfolios, research agencies have come under close scrutiny. Most significantly, Daniels has launched a drive to measure the performance of all branches of the US government, including research and development (R&D) programmes. Those who lobby on behalf of science are alarmed — especially by the idea that spending on fundamental research will be judged by formal 'performance criteria'.

For scientists who have so far remained blissfully ignorant of the OMB, now may be the time to sit up and take notice. The office has become increasingly active in proposing changes to the government's scientific programmes. Early last year, for example, it suggested giving NASA control of the astronomy programmes run by the National Science Foundation (NSF) (see *Nature* **410**, 853; 2001). The idea was later rejected after consideration by a panel of the National Academy of Sciences, but it served as a signal of the OMB's intention to take a more hands-on approach to science policy. And President George W. Bush's second budget request, due to be presented on 4 February, might contain further indications of the OMB's heightened influence.

OMB analysts, who work in the modest New Executive Office Building just a short distance from the White House, are a sharp bunch — in the words of one former employee: "The smartest group of people I've ever worked with."

They need to be — about 200 analysts cover the entire federal budget, with just 20 or so overseeing an annual government R&D spend that currently amounts to some \$104 billion (see graph, opposite).

These bright sparks are typically young, and work punishing hours. One ex-OMB staffer recalls phoning colleagues at home at four in the morning. "There was no griping," he says. Although the professional analysts

are civil servants, and so are not tied to any particular administration, they take their orders from Daniels and the half-dozen other political appointees who run the office from the more opulent surroundings of the Old Executive Office Building, next door to the White House.

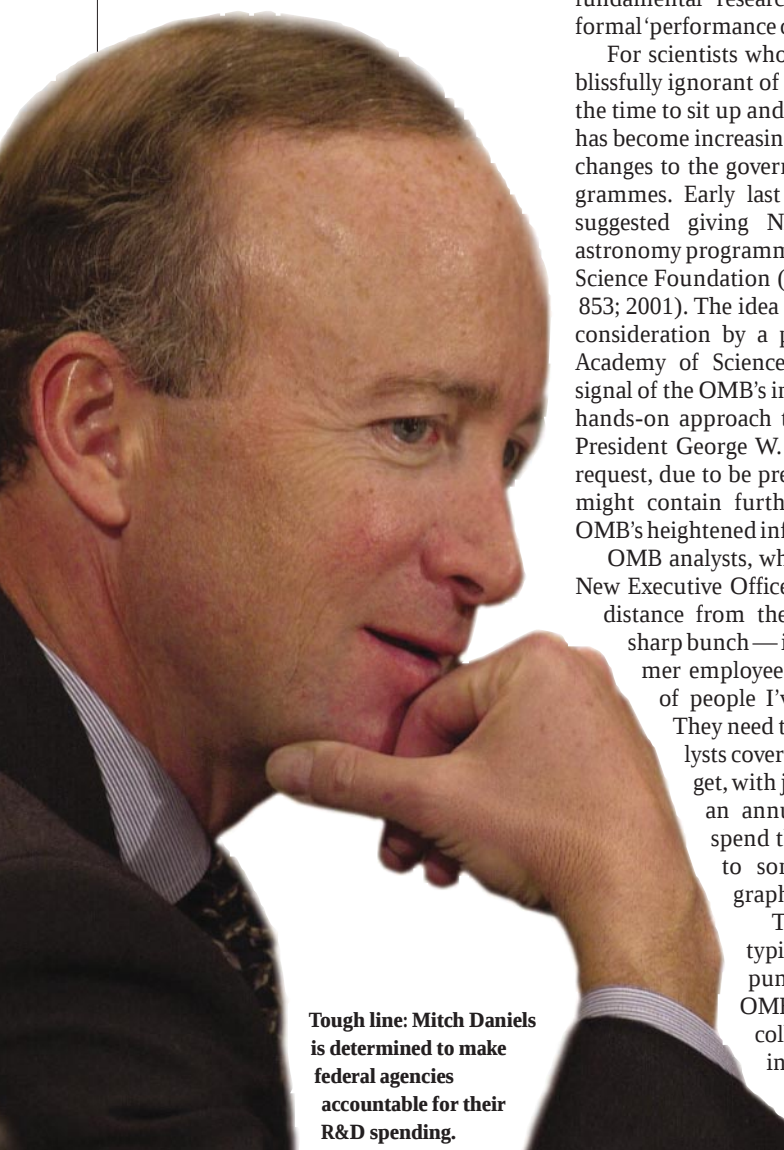
The OMB's current high profile stems from this political inner circle. Daniels, who declined to be interviewed for this article, worked for Republican Senator Richard Lugar (Indiana) and for President Ronald Reagan during the 1980s. In 1997, he became a senior vice-president for the drugs giant Eli Lilly. He now seems to be relishing his public persona as the 'hard man' of the Bush administration. Daniels has been scathingly critical of those — including Republican leaders in Congress — who he sees as being unhelpful to the smooth running of the annual budget process.

Management mindset

The OMB typically exerts more influence when the Republicans control the White House. And the current administration, dominated by a corporate mindset, sets particular store on management efficiency. This isn't simply an aversion to spending money — next week's budget request is likely to propose more deficit spending in a year than Bill Clinton managed in the last four of his presidency. "I think the change is more philosophical," says Nora Noonan, an OMB analyst until 1992 who now runs NASA's National Space Science and Technology Center in Huntsville, Alabama. "It's about being fixated on the president's management agenda."

When it comes to science policy, the OMB's profile has been raised further by Bush's failure to make appointments to the White House Office of Science and Technology Policy (OSTP) until after his science adviser, physicist John Marburger, arrived to head the office in October. "There's been a vacuum in science policy in this administration, and the OMB has filled it," says one congressional official.

Science lobbyists aren't wildly optimistic that Marburger will be able to restore the OSTP's influence. "OMB always calls the shots anyway," says Michael Lubell, a physics professor at the City University of New York and head of public affairs at the American



Tough line: Mitch Daniels is determined to make federal agencies accountable for their R&D spending.

D. COOK/AP



Uprooted? Under a controversial proposal, control of some research, including this Smithsonian Tropical Research Institute study of the forest canopy in Panama, would have moved over to the National Science Foundation.



Physical Society. “The question now is whether the science adviser can influence OMB decision making. I hope Marburger is able to establish himself.”

These worries spilled over on 8 January, when Marburger held a question-and-answer session at the annual meeting of the American Astronomical Society in Washington. Queries about the OSTP itself were few and far between, as Marburger was peppered with anxious enquiries about the OMB’s performance-assessment exercise. “What we need to do is to be more explicit about how we make our choices, and to incorporate that into performance assessment,” Marburger told the restless astronomers.

The performance assessment — tested at parts of the Department of Energy last year and set to sweep through the entire government R&D portfolio this spring — is causing alarm among advocates of basic research. “The science community is worrying about it, as it worries about any threat of accountability — for reasons good and bad,” says an official with a leading university lobby group.

The idea first surfaced in Bush’s management agenda, a 60-page document published in August. Written by the OMB, the document pledged to set “objective investment criteria” for R&D, with those for basic

science scheduled to be agreed and issued this spring (see *Nature* **413**, 5; 2001).

“The guidelines that are being established carry several risks,” says Lubell. In particular, he and other lobbyists are worried at indications that fundamental research will be assessed over a three-year timescale. For Lubell, this is far too short. Another physicist, Mildred Dresselhaus of the Massachusetts Institute of Technology, who ran the Department of Energy’s Office of Science under Clinton, agrees that scientists “have cause for concern”.

OMB officials say that the proposal simply builds on an effort, until recently losing momentum, to implement the 1993 Government Performance and Results Act, which requires every federal agency to measure performance and adjust budgets accordingly. As to the timescale for assessing fundamental research, an OMB official told *Nature* that some agencies were considering reviews of their programmes every three years — but this needn’t mean that research must be measured on its results over this timeframe.

Ratings game

In recent briefings, the OMB has told apprehensive lobbyists and officials that it is considering a list of six criteria to measure research performance. At a packed forum on 15 January, organized in Washington by the American Chemical Society, Marcus Peacock, the OMB’s associate director for natural resources, energy and science, said that the list includes such items as: “Is there a clear public benefit, but a lack of private support?” and “Are there alternatives to direct funding, such as tax incentives?” Ominously, another asks: “Does the programme have an exit strategy?”

The trial run at the Department of Energy got off to a slow start, Peacock admitted to the forum, but he said that the criteria are now shaping up for use in all agencies. The OMB is now trying to reassure scientists about the process. Daniels himself wrote to

Nature in October (see *Nature* **413**, 566; 2001) in an attempt to allay any fears. “We anticipated that people would be afraid that we were going to use exact criteria for basic research, when everyone knows that it is hard to measure or predict its outcomes,” one OMB official told *Nature*, adding that the

final guidelines will be drawn up “with plenty of input from outsiders”. Consultation will include a workshop in late February at the National Academy of Sciences in Washington.

Meanwhile, the OMB continues to instigate policy initiatives aimed, it says, at improving the management of science across government. The proposed transfer of astronomy from the NSF to NASA was one example. Another was the idea, floated late last year, of folding some research programmes of the Smithsonian Institution, the US Geological Survey (USGS) and the National Oceanic and Atmospheric Administration (NOAA) into the NSF.

Critics of the OMB, who are legion, see an ulterior budgetary motive in these proposals. The transfers might, they point out, allow Bush nominally to raise the NSF’s budget, and declare that he was supporting science, while really just moving the cash from other agencies whose involvement in science is less visible. Like the astronomy transfer, the Smithsonian proposal is now reportedly dead (see *Nature* **415**, 252; 2002). But scientists funded by the USGS and NOAA will be scouring the fine print of next week’s budget proposal to see what happens to their programmes.

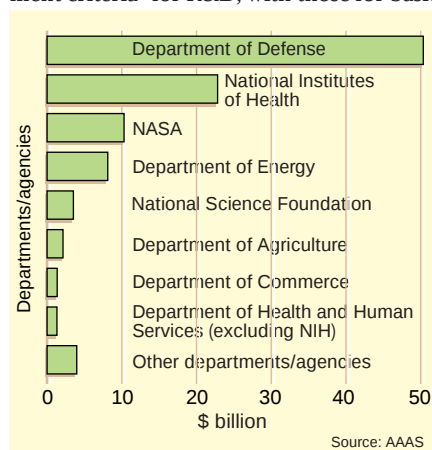
Defenders of the OMB say that it is probing management changes that deserve to be considered. And not everyone sees the OMB’s increased influence on science policy as a bad thing. Professional staff at the office tend to understand the role of basic research better than many other government officials, say the OMB’s supporters. Indeed, many OMB alumni have gone on to fill top administrative posts in research-intensive organizations in both the public and private sectors. “Most scientists have bought into the view that OMB is evil,” says one of its former staff. “The truth is about 180 degrees removed from that.”

But both critics and supporters agree that whether or not the OMB’s individual management proposals are taken up matters little to Daniels and his staff. Although some players in the Bush administration might go to bed fretting that a negative story in *The Washington Post* will shatter their position in the capital’s power structure, OMB staff from Daniels down can sleep soundly in the safe knowledge that their influence will endure. ■

Colin Macilwain is *Nature*’s news editor.

► www.whitehouse.gov/omb

► www.whitehouse.gov/omb/budget/fy2002/mgmt.pdf



Up for grabs: the US government’s spending on R&D will amount to some \$104 billion in 2002.

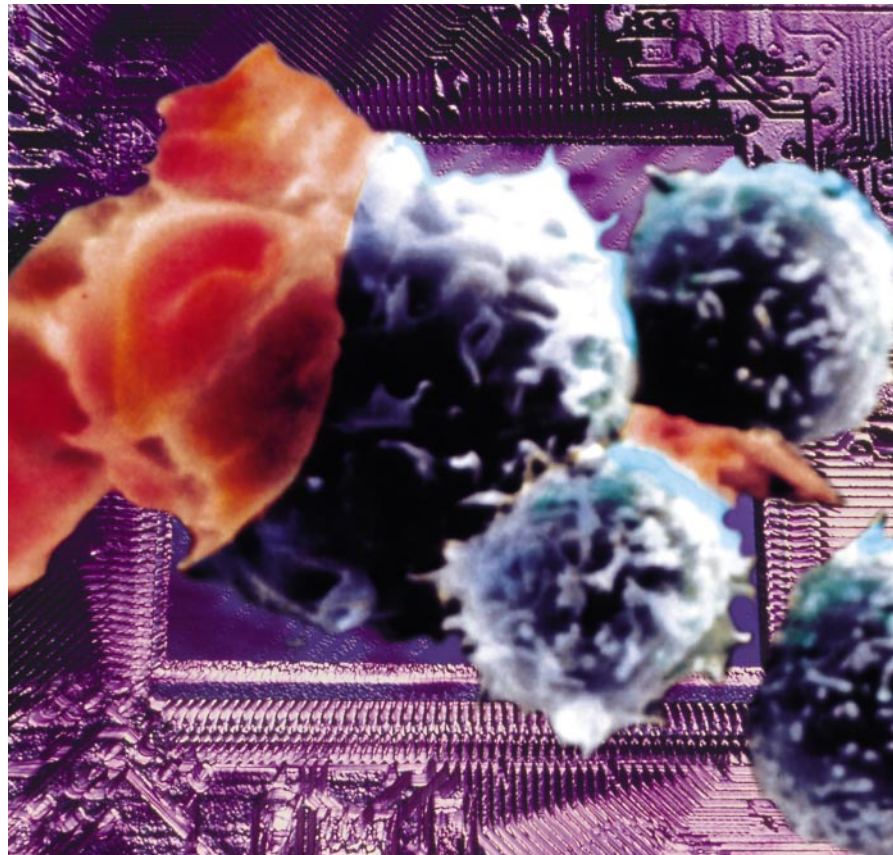
Inspired by immunity

By developing programs that mimic some of the functions of the immune system, computer scientists are tackling problems from fighting fraud to controlling robots. Erica Klarreich investigates.

Computer science has a healthy tradition of stealing nature's good ideas. Programmers have realized that evolution can be simulated using 'genetic algorithms', which drive a computer system towards a problem's most effective solution. The brain has inspired 'neural network' programs that are the basis of many attempts to develop artificial intelligence. But a third biological system — immunity — has until recently attracted little attention.

Computer scientists are now addressing this neglect. In the past few years, they have designed immune-inspired algorithms that might find a wide range of uses, such as detecting fraudulent financial transactions, controlling robots, and even distinguishing cancerous tumours from benign ones.

Our immune system has many desirable attributes. It is versatile and efficient, coping quickly with a diverse array of possible threats. Immunologists estimate that it can respond to



Programmed for success: the functions of white blood cells can easily be translated into algorithms. This montage shows T-cells attacking a large cancer cell (red), superimposed over a computer chip.

10,000 trillion different molecular signatures, or antigens. It is both decentralized and tolerant of errors, so that a few malfunctioning cells, or the loss of part of the system, will not lead to catastrophe. It can also learn to recognize the antigens of specific pathogens and remember them for the future.

Crucially for computer scientists, many of its functions can be translated easily into simple algorithms. "The immune system is rich with inspiration that is mostly untapped at this point," says Jonathan Timmis of the University of Kent in Canterbury, UK, who has devised programs that apply attributes of immunity to pattern-recognition problems.

Although the field is new, several immune-inspired programs are being readied for commercial use. Company 51, a software firm in Redwood City, California, plans to start marketing an immune-inspired system to protect computer networks from hackers and destructive 'worm' programs by the middle of this year. By 2003, Britain's post office, Consignia, could be using an immune-based fraud-detection system to sniff out suspicious transactions.

Fraud detection and computer security are natural candidates for applying the immune system's principles. "There's a pretty direct analogy between the problem of security and the problem the immune system faces in the body," says Stephanie Forrest of the University of New Mexico in Albuquerque, who was one of the first researchers to consider applying the attributes of immunity to computer security in the early 1990s.

Unlike conventional security systems, in which experts try to predict the kinds of

breach that may occur, the immune system starts with little information about the nature of potential invaders. Its ability to respond from a position of ignorance is one of its main attractions for computer-security experts, as the precise forms of computer viruses, hacker attacks or fraud can be hard to predict, and are often only subtly different from benign programs or behaviour.

"Patterns of fraud are often difficult to spot, as a clever fraudster tries to hide his activity, perhaps posing as an employee," says Peter Bentley, who studies immune-inspired security systems at University College London. He notes that pathogenic viruses behave in a similar way, often carrying proteins that mimic those in our own body. "Our immune system has solved this problem, so we'd like to use its processes to solve the problem in fraud detection," he says.

Production line

The immune system constantly reinvents itself, producing 10 million new lymphocytes — white blood cells — every day. The acquired immune system, the part that builds up tailored responses to specific pathogens, has evolved an array of techniques to make these lymphocytes home in on dangerous pathogens. Each lymphocyte either carries receptor proteins that can bind to specific antigens, or produces antibodies with similar specificity.

As lymphocytes are produced, a process of random genetic shuffling results in the receptors or antibodies being put together in subtly different ways, generating a diverse collection of cells that are each specific to a particular



Stephanie Forrest and Peter Bentley draw on the immune system's tricks to fight computer fraud.

antigen. Some receptors may bind to the body's own antigens, so to avoid sending out cells that will trigger such 'autoimmune' attacks, the cells undergo a process called negative selection. In this procedure, each newly made cell is matched up against the body's own molecular signatures, and any lymphocytes that react with them are either destroyed or their activity is suppressed.

The remaining lymphocytes course through the body. Those that fail to interact with an antigen die after a few days. But cells that do interact are stimulated, triggering an immune response. They immediately begin dividing, producing more cells that can detect the same antigen, in a process called clonal selection.

In the case of lymphocytes that produce antibodies, there is a second step which refines the immune response. As cells divide after being stimulated, random mutations alter the antibodies produced. Some of these antibodies will bind to the antigen more effectively — the cells that produce these antibodies subsequently divide more rapidly, and the response becomes increasingly precise.

Lifelong grudge

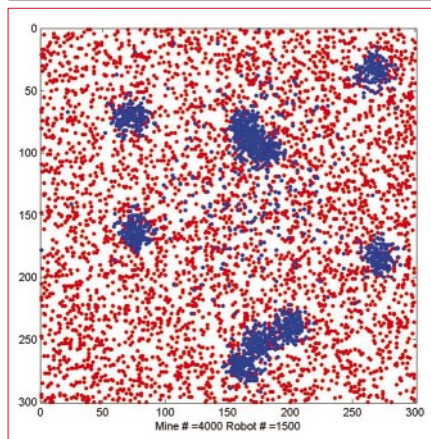
Once the immune system has learned to recognize a particular pathogen, this knowledge is not lost. Some of the cells become 'memory' cells, surviving for long periods and becoming reactivated in response to subsequent attacks by the same pathogen — this is how vaccination works.

Security experts are now trying to mimic these processes. In a typical immune-inspired system, transactions are represented as strings of data, and fraud detectors — equivalent to the immune system's receptors or antibodies — are other strings of data that are compared to entries in the transaction database, or rules that decide whether entries fall into a particular category. One rule, for instance, might recognize any transactions that occur between 2.00 and 2.30 am and involve \$10,000–15,000. Another might be stimulated by transactions in which the buyer enters a London address but a postal code from outside the city.

To build an army of detectors that ignore



Scott Thayer (left) and Surya Singh hope to extend their computer models to real robotics.



This 'minesweeping' model — blue dots are robots and red dots represent mines — uses similar principles to the innate immune system.

normal behaviour but catch fraudulent transactions, the system starts by putting together the components of its rules randomly, just as the immune system creates receptors or antibodies by random genetic shuffling. To weed out detectors that are stimulated by normal transactions, the system then sends the new detectors to 'live' for a period of time in a database of transactions that are known to be legitimate; any detector that recognizes one of these is thrown away.

Once the detectors that recognize normal transactions are removed, the remaining detectors are sent out to monitor the system. Just as lymphocytes that are not stimulated by an antigen die within days, detectors that do not encounter matching transactions are eventually killed off, to make way for rules that seem to match more closely the types of fraud that actually arise.

In a process that is analogous to clonal selection and the refinement of antibody responses, detectors that are alerted by unusual transactions make copies of themselves, adding tiny random mutations that could potentially improve their detective ability. Successful detectors remain in the system as 'memory cells', so that if the same kind of fraud is detected again, the system will respond swiftly. In this way, it learns to recognize common patterns of fraud. If, for example, a certain fraudster typically makes transactions late at night, the system will evolve a large number of detectors that are stimulated by late-night transactions.

Most immune-inspired security systems are similar in basic outline, but all incorporate subtly different attributes of the immune system. "There are just so many wonderful things we've got here," says Richard Overill of King's College London, who leads a team that is working to detect fraudulent post-office transactions. Overill's system, which his team began testing with Consignia this month, concentrates on clonal selection and immunological memory. "Artificial immune systems have the potential to be extremely smart and flexible," he says.

Other researchers are focusing not on the immune system's ability to identify abnormal behaviour, but on its skill in coordinating huge numbers of autonomous cells to create complex responses without a central organizing body. The immune system does this chiefly through local interactions and chemical signals, largely mediated by proteins secreted from its cells. "The immune system has no boss," says Lee Segel, who models immune responses at the Weizmann Institute of Science in Rehovot, Israel. "It has a trillion cells all running around, with no one telling them what to do, but still doing a pretty good job."

This attribute makes the system an ideal model for robotics projects that involve thousands or even millions of small, autonomous robots, argue Scott Thayer and Surya Singh of the Robotics Institute at Carnegie Mellon University in Pittsburgh, Pennsylvania. They are designing robotic systems to clear minefields, carry out search-and-rescue missions, or chart unexplored territory. So far, their work involves computer simulations, rather than real robots — the technology to produce thousands of cheap robots is still under development.

Rapid response

Unlike the fraud-detection researchers, who have focused on acquired immunity, Thayer and Singh have based their program on the 'innate' system, which springs into action at the first moment of an infection. Innate immunity is triggered by receptors that can only recognize pathogens in broad classes, but which respond faster than the more specific receptors and antibodies of the acquired immune system. Usually, they trigger a local inflammatory response, and the lymphocytes involved send out chemical signals to summon other cells.

In Thayer and Singh's minesweeping simulation, virtual robots initially patrol the minefield randomly. When one discovers an area with a concentration of mines, it broadcasts a signal to the other robots, some of which migrate to the area. In the immune system, the strength of secreted chemical signals decreases with distance. To mimic this effect, the virtual robots assign less importance to signals that emanate from far away. This ensures that only those in the vicinity of



Risk-reducing robots: could computerized minesweepers one day replace military experts?

▶ the signalling robot are recruited, while far-away robots continue to monitor other parts of the minefield.

This simple mechanism easily outperforms traditional 'raster' scans, which divide a search area into a grid and comb it cell-by-cell, row-by-row. Thayer and Singh's model is particularly effective for 'moving' minefields, in which the mines can jump, foiling raster scans by moving into previously cleared areas.

Thayer and Singh are working to incorporate the acquired immune system's ability to learn into their program. Although robots cannot copy themselves, if a robot finds itself to be especially well adapted to defusing a particular type of mine, it could broadcast the details of its program to nearby robots, duplicating its successful software, rather than its hardware.

Mutant minesweepers

The researchers also want to include mutation, allowing robots to change randomly the way they carry out operations, or the relative importance they assign to different signals; successful mutations could then be broadcast to others. For instance, a mutation might make a robot check a mine twice before summoning its neighbours. "As with the immune system, because there are so many robots, you can afford to risk two or three in trying out new behaviour," says Singh.

Computer scientists are even appropriating outdated immunological theories. In 1974, Niels Jerne of the Basel Institute for Immunology in Switzerland, who later won a Nobel prize for his theoretical work, introduced the idea of 'immune networks' to try to explain how the immune system maintains its memory of an antigen. Jerne proposed that certain antibodies, termed 'anti-antibodies', bind to other antibodies, while 'anti-anti-antibodies' recognize these, and so on. All of these antibodies, Jerne argued, suppress or stimulate one another in a way that is held in equilibrium until the original antigen comes along again and disturbs the network, triggering an immune response.

The theory has since fallen from favour — although antibodies that recognize other antibodies do exist, there is no strong evidence that they function in self-regulating immune networks as Jerne proposed. Nevertheless, computer scientists have taken inspiration from the idea to develop systems for discovering patterns in vast amounts of data. "From a computer-science point of view it's a lovely theory," says Timmis.

He is developing a pattern-recognition program that could carry out tasks such as finding relationships in people's Internet shopping habits or sifting through epidemiological data to identify risk factors for disease. His data-mining program uses virtual antibodies that cluster into networks, much as Jerne suggested. The program takes a small, random sample of a data set as its

starting 'antibody' population; the full data set is the collection of 'antigens'. If the program is looking for patterns in Internet shopping habits, each antibody and antigen might be a consumer profile that consists of a list of, say, 100 products together with details of whether or not the consumer bought each.

First, the antibodies are compared to each other, and ones that match closely are linked together to form a network. Two different profiles might be considered to match if they agree in, say, at least 50 of the entries. This matching threshold is set before the program starts running, and the level chosen affects the strength of the patterns found.

Once the antibodies have been linked into a network, they are compared to the antigens in the full data set. Antibodies that are not similar to enough antigens are removed, whereas antibodies that match many antigens make copies of themselves, introducing small mutations as they replicate. These mutated copies are then incorporated into the network, linked to similar antibodies. The entire process is repeated many times so that the network evolves clusters, each of which represents some pattern in the original data set.

If, for example, Internet shoppers who buy books online also tend to buy music, the original set of antibodies might contain a few consumer profiles that include both books and music. As these profiles are similar to each other, they will be linked together in the network, and as they are also similar to many profiles in the full data set of antigens, every time they are compared to the antigens they will be stimulated to make mutated copies of themselves. These copies are then linked in the network to the original profiles that include books and music, and gradually a cluster of antibodies forms. "Often the clusters will find obvious patterns that you knew already, but sometimes you get real nuggets of information," says Timmis.

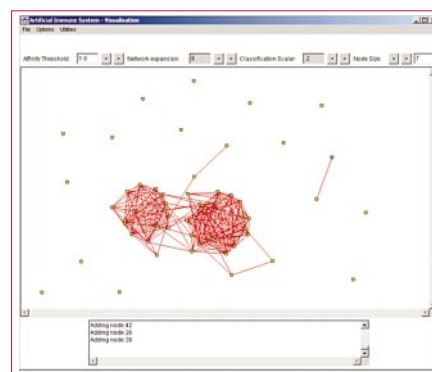
Breast test

Timmis has tested his algorithm on the Wisconsin Breast Cancer Database, which contains patient profiles and data from their breast-tumour biopsies — documenting the sizes of clumps of cells, uniformity of cell sizes and shapes, and so on. The program learned to distinguish between samples from malignant and benign tumours.

Immune networks may prove to be more flexible learning tools than brain-inspired neural networks, which learn patterns by changing the strength of connections between simulated neurons. "With a standard neural network you have to define the network structure first, and the network is only as good as its initial design," says Timmis. By contrast, immune networks allow their components to change, as well as the connections between individual components, giving added plasticity.



Jonathan Timmis's 'immune network' program identifies patterns in huge data sets by evolving clusters of successful antibodies. This enables it, for example, to distinguish tumour types (below).



Researchers working in the field expect to garner many more computational insights from mimicking the immune system — particularly given that our biological understanding is advancing all the time. Forrest notes that immunologists have recently become very interested in the innate system. "We modellers are far behind," she says.

Some biologists argue that computer scientists may also provide insights for immunologists. "I think it would be beneficial to see more interplay," says Alan Perelson, a mathematical biologist who studies problems in immunology at the Los Alamos National Laboratory in New Mexico. "If the computer-science people have to go outside of 'normal' immunology, they might find something we've missed."

In the meantime, computer scientists have plenty of work to do in exploiting well-understood attributes of the immune system. "Nature has been perfecting this for so many years that even if we get it just one-tenth right, it's often better than what we could come up with on our own," says Bentley.

Erica Klarreich has just finished an internship with Nature.

Fraud detection

♦ www.dti-mi.org.uk/newweb/cifd.html

♦ www.icsa.ac.uk/Projects/link.html

Robotics

♦ www.ri.cmu.edu/pubs/pub_3845.html

Pattern recognition/data mining

♦ www.cs.ukc.ac.uk/people/staff/jt6

JONATHAN TIMMIS

False samples are not the same as blind controls

Informal efforts to “test” a laboratory corrupt the data stream, where integrity is crucial.

Sir — As director of the laboratory that analysed the samples for the National Lynx Survey sponsored by the United States Department of Agriculture's Forest Service (USFS), I am responding to the damaging yet unsubstantiated conclusions in your Opinion article “Lynch mob turns on lynx researchers”¹ and News article “Fur flies over lynx survey's suspect samples”² about sampling for Canada lynx (*Lynx canadensis*).

My laboratory developed the protocol to distinguish among species of North American felids based on DNA analysis of hairs using known individuals across the geographical range of the target species, and comprehensively validated it using blind control tests both within our lab and at another lab. Our protocol was peer-reviewed and published³.

In practice, positive and negative controls are used with every sample set analysed. We have never misidentified a species when the identity of the sample was known, and we consistently detect lynx in areas where they are known to exist. In contrast to the implications of your Opinion article, therefore, the laboratory procedure to distinguish lynx based on hair samples is fully diagnostic and validated.

In 1999, the USFS implemented the National Lynx Survey to evaluate systematically the distribution of lynx on federal lands across 12 states; species identification based on DNA analysis was the final step. I was principal investigator in charge of DNA analysis, and Kevin McKelvey from the Rocky Mountain Research Station (USFS) was the principal investigator for the nationwide survey protocol. Field personnel (several hundred permanent and temporary federal and state employees) collected hairs from over 13,000 rub pads⁴ and sent them to the lab in labelled vials with accompanying data sheets. Because integrity of the data stream is particularly crucial in a study of such enormous scope and scale, the protocol was explicit, with comprehensive written instructions for all aspects of gathering, labelling and submitting samples.

For a field worker to arbitrarily decide “to test the lab” by labelling a hair from elsewhere as if it were a field-collected sample corrupts the integrity of the data and does not constitute a blind control. By analogy, a medical field worker who surreptitiously contaminates a blood sample to “test the rigour” of the serological protocol during a wide-scale survey to assay a disease is not conducting

a blind control but is fabricating data that could lead to false conclusions about disease distribution.

I do not know the field personnel who submitted false samples and cannot assume anything about their motivation. But the vast majority of those participating in this unprecedented large-scale survey followed scientific protocol. The few who misrepresented data without notifying those who analyse and interpret the data did not, as your Opinion article suggested, merely implement blind controls.

Nobody contacted either me or the USFS research scientist in charge of the survey protocol about any concerns or told us about the plan to misrepresent data. In fact, we learned about the 1999 false samples only after the start of an internal USFS investigation.

I agree with your Opinion article that “Clean data are needed to prevent years of court battles over use of the forests”¹. In

this case, the methodology was peer-reviewed, published, explicit, complete and followed by all except those few who intentionally mislabelled samples. The only potential smudge on the ‘cleanliness’ of data in the National Lynx Survey at this point comes from submission of samples not in compliance with scientific protocol.

Our Letter published on page 424 of this issue⁵ is not part of the National Lynx survey; all 599 samples used in ref. 5 were collected from freezers and other known sources.

L. Scott Mills

Wildlife Biology Program, School of Forestry,
University of Montana, Missoula,
Montana 59812, USA

1. *Nature* **415**, 101 (2002).
2. Dalton, R. *Nature* **415**, 107 (2002).
3. Mills, L. S., Pilgrim, K. L., Schwartz, M. K. & McKelvey, K. *Conserv. Genet.* **1**, 285–288 (2000).
4. McDaniel, G. W., McKelvey, K. S., Squires, J. R. & Ruggiero, L. F. *Wildl. Soc. Bull.* **28**, 119–123 (2000).
5. Schwartz, M. K. *et al.* *Nature* **415**, 424–426 (2002).

Mislabelling muddies the forest-survey waters

Sir — Calling people who object to deliberate mislabelling of samples a “lynch mob”, as your Opinion article (*Nature* **415**, 101; 2002) does, is like asserting that the collapse of the US energy company Enron was lamentable, and that those employees who lost their retirement savings deserve to be ashamed. It is half true (it is lamentable), but the implied criticism is completely misdirected.

Although conservatives in Congress and the media have indeed harvested a bounty of partisan hay from the misbegotten actions of some agency employees, several important issues remain unresolved.

It is not clear whether the field workers had any fraudulent intent or whether they intended the submitted lynx (*Lynx canadensis*) hairs to be controls in a test of the analytical lab. But it is clear that earlier samples submitted from the same geographical region to a different lab produced what were subsequently believed to be false positives — lynx hairs were reported from areas with no previous or subsequent evidence of lynx.

How the submission of blind positive controls could have provided a check on false positives is far from clear. More important, the laboratory supposedly being tested in the episode described in

your article was in no way suspect in the earlier problems.

Any criticism should be levelled at the field workers who undertook a series of undocumented, informal steps that could be misinterpreted as fraud. Clearly, all would have benefited had they written a protocol, received documented approval, and notified the analytical lab that blind controls would be included in submitted samples.

Steven W. Buskirk

Department of Zoology and Physiology, University
of Wyoming, Laramie, Wyoming 82071-3166, USA

Regional network raises profile of local journals

Sir — Researchers worldwide strive to publish in journals with high impact factors, but such journals are concentrated in developed countries, leaving journals in developing countries ranked at the bottom (see S. B. Vohra & D. Vohra, *Nature* **412**, 583; 2001). These journals are further imperilled by the decreasing level of support among local subscribers, who in the face of economic constraints may prefer to subscribe to journals with high impact factors.

Electronic publishing is bringing hope to these threatened journals. Citations increase when papers are freely available on

the Internet (see S. Harnad, *Nature* **410**, 1024–1025; 2001), and journals can increase their impact factors by publishing their contents electronically (see M. Curti, V. Pistotti, G. Gabutti & C. Klersy, *Haematologica* **86**, 1015–1020; 2001).

However, two shortcomings still stand out. First, online publication involves substantial costs (software development, hosting support, and so on), despite potential price cuts associated with the reduction of print editions. Second, free online access to local journals does not necessarily lead to increasing readership without a powerful means of dissemination. For these journals to be internationally recognized, regional networks with speedy access from search engines, portals and indexing services are required.

Such a network was launched in 1997: SciELO (Scientific Electronic Library Online: www.scielo.org) is a publicly funded initiative set up to promote cooperative, free electronic publishing of scientific journals from developing countries; the development of regional databases; and the implementation of indicators of scientific literature usage. It currently comprises 93 journals from Brazil, Chile and Cuba. We assessed its international impact by comparing the impact factors of journals before and after being incorporated in SciELO.

We found five Brazilian journals that had been indexed by ISI for at least five years and available in SciELO for at least two. The impact factors of these journals more than doubled (132.7% increase, one-tailed Wilcoxon signed ranks test, $P < 0.02$) since their inclusion in SciELO.

This indicates that such networks not only foster the availability of scientific information on a regional scale, but also generate international impact which may entice researchers into publishing in the journals concerned. Those who fund and promote regionally coordinated networks for scientific electronic publishing can help developing countries to revitalize their local journals and enhance the international representation of locally generated knowledge.

Wladimir J. Alonso, Esteban Fernández-Juric

Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

Laboratories' gravity train has ground to a halt

Sir — Your Opinion article “Time to halt the gravity train” (*Nature* **414**, 829; 2001) questions the salary and benefits packages paid by international laboratories. It is

entirely correct that these should be scrutinized at a time when cost efficiency is high on the agenda of most institutions, but some of the comparisons you make are oversimplified and misleading. The reality is that the gravity train ground to a halt long ago.

In comparing CERN (the European laboratory for particle physics) with DESY, Germany's main high-energy physics laboratory, which is in Hamburg, for example, you fail to take into account the significant cost-of-living differentials between Geneva and most other European locations.

As you rightly point out, CERN conducts regular tracking studies which compare our salary and benefits packages with industry and other institutions that compete for the scientific, engineering, IT and other skills we need. In this context, CERN is primarily a provider of research infrastructure, supporting a user community of more than 6,000 scientists — only 90 of our 2,700 current staff are engaged in fundamental research. We are far from topping the compensation table: indeed, in some skills areas we find it increasingly difficult to attract suitably qualified people away from their home markets, particularly in northern Europe.

As your leader implied, multinational institutions must be increasingly efficient in managing their resources and in being able to justify the investments they make. Maintaining the skills levels and human vitality of CERN is fundamental to being able to fulfil its mission as one of the world's leading scientific research establishments. In my view, the budgetary pressures we and many comparable institutions face are such that one of our biggest challenges will be to remain competitive in this respect.

Luciano Maiani

Director general, CERN, CH-1211, Geneva 23, Switzerland

Tax-free pay lets funders evade responsibilities

Sir — Although I agree with your Opinion article (*Nature* **414**, 829; 2001) that the tax-free salary status of a few scientific researchers in Europe is an anachronism, I dispute your use of the phrase “gravity train”. Many life-science researchers are paid apparently generous tax-free stipends at some point in their careers, but these salaries are less generous after making reasonable provision for healthcare and retirement. Until Europe-wide agreements on these standard employment benefits have been reached, the alleged “gravity

train” will continue.

You do not mention that government organizations such as the Max Planck Society in Germany often pay postdoctoral researchers tax-free stipends that can hardly be described as generous — between 17,850 euros (US\$15,344) and 22,000 euros a year, depending on the exact circumstances. After private health-care (about 2,500 euros a year) and a basic pension (about 2,000 euros) have been deducted, this is not a great deal of money.

This practice should be stamped out. Surely scientists deserve the same pay and working conditions as other professionals? The era of the tax-free stipend is indeed over, but for very different reasons from the ones you suggest.

It is easy to print melodramatic headlines about the high tax-free salaries of a few physicists at CERN and to overlook the real problems faced by many more scientists around Europe. Tax-free stipends are simply a way that scientific institutions and funding bodies escape responsibilities to their employees.

Francis Barr

Department of Cell Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18a, Martinsried 82152, Germany

People, payments and positions at DESY

Sir — In your Opinion article about benefits for scientists working abroad (*Nature* **414**, 829; 2001) you state that only one-third of the scientists at DESY (Germany's high-energy accelerator centre) are German.

This is correct for the scientists who visit DESY to perform experiments as members of one of the four HERA collaborations or in the Synchrotron Radiation Laboratory HASYLAB. They are, however, not employed by DESY. Of the scientists on the DESY payroll, 75.4% are German.

You also state that at DESY a young, unmarried graduate can expect to earn 18,500 euros (US\$16,346) a year. That figure is incorrect. The correct amount is 40,000 euros.

Petra Folkerts

DESY, Notkestrasse 85, D - 22607 Hamburg, Germany

The “young, unmarried graduate” positions discussed in the Opinion article referred to PhD students, who in Germany are usually paid a half-salary, which is typically 18,500 euros. People earning the full-time salary of around 40,000 euros cannot study for a PhD. — Editor, Correspondence.

Mad hatters at the DNA tea party

A unique discovery brings fresh insight to the discovery of DNA's structure.

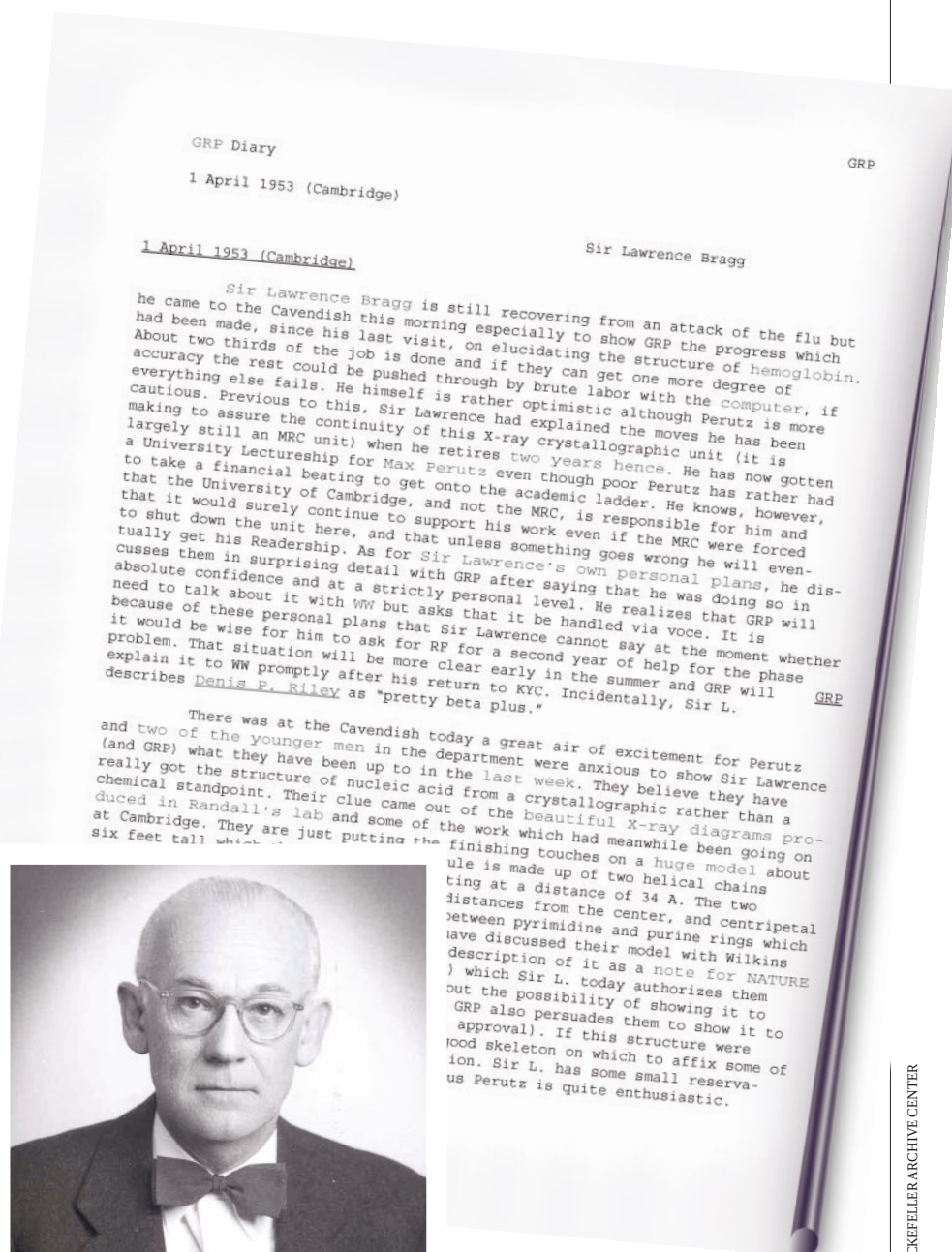
Jan A. Witkowski

The only contemporary account by an observer of events surrounding the discovery of the double-helical structure of DNA has just been unearthed in the Rockefeller archives. The memorandum (reprinted in facsimile here) is from the diary of Gerard Roland Pomerat ("GRP"), assistant director of the natural sciences programme at the Rockefeller Foundation. Pomerat was at the Cavendish Laboratory in Cambridge visiting its director, Sir Lawrence Bragg, whose research on protein structure was being funded by the Rockefeller. Pomerat's director, Warren Weaver ("WW"), was the first to use the phrase "molecular biology", having been, as Robert Kohler put it, "a midwife at what was later seen to be its birth".

Bragg and his father, William Bragg, were pioneers of the application of X-ray crystallography to organic and biological macromolecules. The Cavendish Laboratory had already made a major contribution to the analysis of protein structure before Bragg's arrival in 1938; in 1934, J. D. Bernal obtained the first X-ray-diffraction picture of a protein, the enzyme pepsin. Bernal moved to Birkbeck College, London, but one of his group — Max Perutz — remained in Cambridge, working on the structure of haemoglobin. Although at first not much interested in Perutz's results, Bragg became enthusiastic and arranged funding for him from the Rockefeller. Later, in 1947, Perutz was a founding member of the Medical Research Council's (MRC's) unit for the study of the "molecular structure of biological systems" within the Cavendish Laboratory.

But at the time of Pomerat's visit, the future of the MRC unit was uncertain: Bragg was to retire in 1955 and his successor might not want a biology group in an overcrowded physics laboratory. Bragg had persuaded the university to make Perutz a tenured lecturer, and he himself was considering moving to the Royal Institution. Bragg's discussions with Pomerat had to be "in absolute confidence" because Bragg was in an embarrassing position. He had probably supported the removal of the Royal Institution's previous director, and as he was now likely to be offered the post, he wanted to avoid giving the incorrect impression that he had been scheming to create a job for himself.

Pomerat had learned of Perutz's progress, which was, in fact, very slow, despite some



Gerard Roland Pomerat witnessed the "great air of excitement" at the Cavendish Laboratory in April 1953.

technical advances. For example, Pomerat's comment on "the computer" refers to Hugh Huxley's use of the university's Electronic Delay Storage Automatic Calculator (EDSAC), to carry out, in 30 minutes, a calculation of a Patterson projection that had previously taken two weeks. The most exciting research was from "two of the

younger men", Jim Watson and Francis Crick. Pomerat contrasts their careers. While Watson had been moving in the *avant garde* circle of phage genetics centred on Max Delbrück, Crick's career as a physicist had been interrupted by the Second World War, and his move into biology had become bogged down in uninteresting biophysical

There was at the Cavendish today a great air of excitement for Perutz and two of the younger men in the department were anxious to show Sir Lawrence (and GRP) what they have been up to in the last week. They believe they have really got the structure of nucleic acid from a crystallographic rather than a chemical standpoint. Their clue came out of the beautiful X-ray diagrams produced in Randall's lab and some of the work which had meanwhile been going on at Cambridge. They are just putting the finishing touches on a huge model about six feet tall which shows that the molecule is made up of two helical chains running parallel to each other and repeating at a distance of 34 Å. The two coils run one above the other, at equal distances from the center, and centripetal to them are a series of direct linkages between pyrimidine and purine rings which in turn, are linked to the helices. They have discussed their model with Wilkins of Randall's group and have written up a description of it as a note for NATURE (which would probably appear in early May) which Sir L. today authorizes them to send. They are particularly excited about the possibility of showing it to Pauling, who will be here this week, but GRP also persuades them to show it to Todd first (and to this Sir L. gives full approval). If this structure were the correct one it would present a very good skeleton on which to affix some of the modern theories of chromosome duplication. Sir L. has some small reservations about it all, but the usually cautious Perutz is quite enthusiastic.

research with Arthur Hughes at the Strangeways Laboratory in Cambridge. Crick wrote later that the best thing about this project was that it left him plenty of time to improve his knowledge of biology. He was now working for his PhD under Perutz's direction, but, to Bragg's chagrin, he was apt to spend more time reflecting on other people's projects than his own. It was not until 1953 that Crick completed his PhD.

Pomerat refers to other researchers. John Kendrew had begun his analysis of sperm-whale myoglobin in 1948 and nine years later determined its structure at 6-Å resolution. Perutz completed the structure of haemoglobin in 1959, for which he and Kendrew shared the 1962 Nobel prize for chemistry. Huxley went on to achieve great success with his structural studies of muscle fibres, whereas Vernon Ingram used Fred Sanger's peptide-fingerprinting technique to determine the difference between normal and sickle-cell haemoglobin.

Maurice Wilkins had begun his X-ray analyses of DNA in 1950 in John Randall's MRC Biophysics Unit at King's College, London, but the "beautiful X-ray diagrams" mentioned by Pomerat were taken by Rosalind Franklin. Randall had suggested before she joined the unit that she should work on proteins in solution, but later, just before Franklin arrived, proposed instead that she work with Ray Gosling, extending the initial findings on DNA. Randall wrote that "as far as the experimental X-ray efforts are concerned there will be at the moment, only yourself and Gosling". Unfortunately, Wilkins was not told of this change of plans, and he was away on holiday when Franklin arrived at King's in January 1951. He returned to find that his project had

seemingly been hijacked, leading to much ill will. Franklin's major contribution was to distinguish clearly the A and B forms of DNA. In May 1952, she and Gosling took a photograph of the B form that later set Watson's pulse racing and convinced him and Crick that DNA is helical.

Bragg insisted that Alexander (later Lord) Todd inspect Crick and Watson's model for its chemical accuracy, because in 1950 the Cambridge group had published a paper that Bragg later called "the most ill-planned and abortive in which I have ever been involved". They had not realized that the peptide bond is planar and thus failed to discover the α -helix, described by Linus Pauling and Robert Corey in 1951. Todd later wrote that he had asked Bragg: "Don't you take any chemical advice when you do this kind of work?" — hence Bragg's insistence that Todd view the DNA model.

In *The Double Helix*, Watson cast Pauling as the deadly competitor, but it seems that Pauling regarded DNA simply as another large molecule to be solved. Pauling visited the Cavendish on Friday 6 April 1953 and, according to Watson, "gracefully, he gave his opinion that we had the answer".

Elizabeth, Watson's sister, had typed the final version of the famous *Nature* paper on Saturday 28 March, and it was this manuscript that Pomerat tells us "Sir L. today authorizes them to send". Watson says that the manuscript was mailed on Thursday 2 April; a later letter from Wilkins tells Crick that "all is safe in the hands of Gale (editor of *Nature*) who may be able to get it in April 25". And indeed the three papers — one from Cambridge and two from King's — appeared in the issue of Friday 25 April 1953.

The DNA story is so firmly centred on the main players that it is easy to forget that there were others working in the field. Pomerat mentions Denis P. Riley, who carried out X-ray and light-scattering studies of colloids, and R. D. Fraser, a colleague of Wilkins, who developed a triple-helix model of DNA. Wilkins had wanted the latter published along with the double-helix papers, but Watson and Crick demurred; the only surviving reference to Fraser's model is a single sentence in the first *Nature* paper.

Watson tells us in *The Double Helix* that Francis Crick's excitement grew each day as he regaled visitors with the wonders of the structure. Now, Pomerat's words give us a new view of what the visitors saw of Watson and Crick: "Somewhat mad hatters who bubble over about their new structure." ■

Jan A. Witkowski is at the Banbury Center, Cold Spring Harbor Laboratory, PO Box 534, Cold Spring Harbor, New York 11724-0534, USA.

Acknowledgements

I thank David Micklos, who found the Pomerat memorandum; Darwin Stapleton of the Rockefeller Archives for permission to reprint it; and Robert Olby for comments.

Further reading

- Crick, F. H. C. *What Mad Pursuit: A Personal View of Scientific Discovery* (Basic Books, New York, 1988).
- Judson, H. J. *The Eighth Day of Creation: The Makers of the Revolution in Biology* (Cold Spring Harbor Laboratory Press, New York, 1996).
- Kohler, R. E. *Partners in Science: Foundations and Natural Scientists 1900–1945* (Univ. Chicago Press, 1991).
- Olby, R. *The Path to the Double Helix* (Dover, New York, 1994).
- Swann, B. & Aprahamian, F. J. D. *Bernal: A Life in Science and Politics* (Verso, London, 1999).
- Thomas, J. M. & Phillips, D. *The Legacy of Sir Lawrence Bragg* (Royal Institution of Great Britain, London, 1990).
- Watson, J. D. *The Double Helix* (Atheneum, New York, 1968).

A certain chemistry

Neurologist Oliver Sacks recounts a childhood love affair.

Uncle Tungsten: Memories of a Chemical Boyhood

by Oliver Sacks

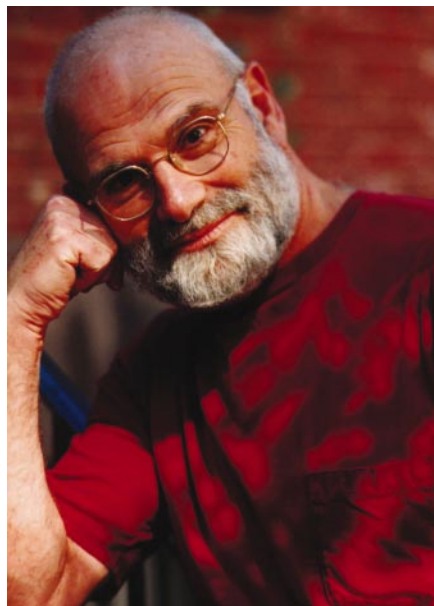
Picador/Knopf: 2001. 340 pp. £17.99/\$25

John Emsley

How odd. Why should Oliver Sacks, a world-renowned neurologist, and author of many influential and popular books, such as *The Man Who Mistook His Wife for a Hat*, write a book about chemistry? The reason is that he fell in love with the subject as a boy, and by embracing it he could find refuge from the storms of life that raged around him.

My first reaction to *Uncle Tungsten* was to see parallels with Primo Levi's *The Periodic Table*, in which Levi recounted life in Auschwitz, and where his love of chemistry also sheltered him — and ensured his survival, for his chemical training was useful to the Nazi war effort. Both books cover the same period of history, although they tell of very different lives. And through them both runs a passion for a science that has now become unpopular. When Sacks was young, chemistry delivered a cornucopia of benefits that changed our homes, our health, our food supply, our leisure, and even the way we looked and smelt. No wonder that he, like many of his contemporaries, was eager to explore the world of the test tube.

Sacks was born in London in 1933, and this book is an autobiography of his schoolboy years. His Jewish parents, who were busy doctors, encouraged his hobby of chemistry, but it was his Uncle Dave who demonstrated



Sacks's schoolboy infatuation with chemistry shielded him during turbulent times.

its practical use at first hand. He ran Tungstallite, a firm that made the tungsten filaments for electric light bulbs, and he is the uncle of *Uncle Tungsten*. Sacks's parents provided space for their son to equip a home laboratory and, although he was only a boy, they let him sally forth to local suppliers of chemicals and purchase anything that he needed for his experiments. Today it seems unimaginable that highly flammable solvents, corrosive acids, toxic salts and potentially explosive

The young Sacks, seen here as a baby with his family, was encouraged in his hobby by his parents, but it was his Uncle Dave (inset) who demonstrated its practical use.



materials could be bought so easily.

However, *Uncle Tungsten* is not about the simple experiments Sacks was able to carry out, but is a scholarly work about chemistry itself. He delves into the history of the subject, the personalities of the early chemists, such as Lavoisier, Dalton and Davy, the nineteenth-century developments that led to the periodic table of Mendeleev, and thence on to radioactivity, the nature of the atom and the development of modern theories. The chapter on Mendeleev is a particular delight, confirming that Sacks has not only read about the great Russian chemist's achievements but has researched those of Mendeleev's contemporaries who came near to discovering the periodic table as well.

Running through the book we have the equally fascinating story of a young Jewish boy who was evacuated from London to escape the Blitz. Indeed, Sacks's life might well have ended early when a large bomb fell in the garden next door; luckily, it failed to explode. Sacks and his older brother Michael were packed off to the apparent safety of boarding-school in the Midlands. But what protected them from the dangers of war exposed them to an emotional blitz of bullying. Indeed, so searing was the experience that Michael became mentally ill.

Sacks returned to London in 1943. But that city's tribulations were far from over. During the next 18 months it experienced the baby-blitz, the flying bombs and the V2 rockets. However, Sacks barely mentions them, and this gives us some idea of the intensity of his love affair with chemistry. The hardships and deprivations of life, and the goings-on of his eccentric family, could be ignored, as long as young Sacks could retreat into his chemical world in Cricklewood.

One of the hardest tribulations he faced as a boy is described in the chapter "Ma". There, he tells how his mother tried to encourage him to take up a medical career by having him carry out an autopsy. The young 14-year-old Sacks was given the body of a 14-year-old girl to dissect! How welcome, then, must have been the emotional safety of his preferred science. But within a year or two, his love affair with chemistry was ending. Yet, clearly, it was never completely over. For another 50 years Sacks continued to carry a candle for his old flame, and now we can enjoy a tale that fuses autobiography and chemistry, to cast a spell that kept me turning the pages of one of the most enjoyable books I have read for many years. ■

John Emsley is science writer in residence in the Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

A heated fight against devastation

Forest Fires: Behavior and Ecological Effects

edited by Edward A. Johnson

& Kiyoko Miyanishi

Academic: 2001. 594 pp. \$74.95

Daniel Nepstad

Major forest fire research programmes were set up during the twentieth century in an attempt to reduce the considerable costs of forest fires to society and to provide a scientific basis for forest fire management systems. A new synthesis of forest fire science that promises to distil the vast scientific literature produced by these programmes is a welcome contribution. Edward Johnson and Kiyoko Miyanishi have edited this volume to foster a more mechanistic, physically based approach to forest fire science, going beyond the 'traditional' descriptive approaches. The book is intended to bridge the gap between elementary forest fire texts and the highly technical literature on the subject. To these ends, it is partially successful. The book provides a rich summary of our current knowledge of several important aspects of forest fire science, from fuel dynamics to coupled atmosphere–fire modelling. But it lacks a synthesis of the ecological and economic importance of forest fires and an integrated analysis of the critical gaps in our understanding of this phenomenon.

Dozens of careers and decades of research have been devoted to forest fire science in the United States, Canada, Australia and Russia. But in the emerging forest fire hotbeds of Brazil, Indonesia and Mexico, forest fire scientists can be counted on the fingers of one hand. These disparate scientific communities have very different reading needs. The decades-old forest fire school has already developed and tested empirically based approaches to assessing the risks of forest fires. It has also modelled fire behaviour and fuel dynamics in many forest types, and documented the ecological effects of forest fires. Fire laboratories in these countries examine the physical and chemical aspects of forest fires under controlled conditions. Johnson and Miyanishi target this school, hoping to speed up its transition to physically based forest fire science. And it is for this audience that the book will be most useful.

The incipient forest fire research communities of developing nations are just beginning to conduct the type of descriptive research that Johnson and Miyanishi hope to move beyond. We still don't know how much of the Amazon or Bornean rainforests catch fire each year, for example; at least 10,000 square kilometres of standing forest were ignited in each region during the severe



Burning issue: research into forest fires has escalated over the past decades.

droughts of the 1997–98 El Niño episode. The book highlights a pioneering study of forest fires in Amazonia by Christopher Uhl and Boone Kauffman as an example of the flawed, descriptive forest fire research that must be avoided. But Uhl and Kauffman's is both an important and a timely contribution, documenting effects of selective logging on the mass and moisture content of the fuel layer. The way in which the leaf canopy regulates the flux of radiant energy to the forest interior is the single most important determinant of tropical rainforest flammability, but this is not mentioned in *Forest Fires*.

An effective bridging of the gap between elementary forest fire texts and the highly technical forest fire journals requires overview chapters describing the ecological and economic importance of forest fires on a global scale, and indicating the most important gaps in the science. The absence of such chapters in *Forest Fires* means that the forest fire neophyte encounters chapters rich in equations and literature references, but poor in didactic explanations of the basic principles of forest fire behaviour and its ecological effects. An issue as important as the effect of global warming on future fire regimes is tucked away in a chapter on "Climate, weather and area burned", illustrated by illegible maps.

The book's strength is as a state-of-the-art review of research on pyrolysis, flames, lightening, fuel–moisture dynamics, smoke, combustion chemistry, and more. Every chapter carefully defines its notation, and some of the chapters are written in engaging prose. But the editors' odd attack on descriptive research sends the wrong message to prospective forest fire scientists in regions suffering rapidly escalating problems with

such fires. The first step in establishing research programmes in places such as Amazonia, Borneo and Mexico is to walk through burning and burned forests, talk to property owners who ignite and fight fire, and quantify and systematize these field observations. My fire science students and colleagues in the Brazilian Amazon use *Forest Fires* as a reference. But the inspiration for their studies is still the classic *Introduction to Wildland Fire* by Stephen Pyne, Patricia Andrews and Richard Laven (Wiley, 1996). ■

Daniel Nepstad is at the Woods Hole Research Center, Woods Hole, Massachusetts 02543, USA, and the Instituto de Pesquisa Ambiental da Amazônia, CEP 66087-430, Belém, Pará, Brazil.

Every second counts

The Measurement of Time: Time, Frequency and the Atomic Clock

by Claude Audoin & Bernard Guinot

(translated by Stephen Lyle)

Cambridge University Press: 2001. 346 pp.

£75, \$110 (hbk), £27.95, \$40 (pbk)

Ken Johnston

In our daily lives, time regulates our activities. Almost everyone wears a watch. The accuracy with which time is measured increased by orders of magnitude in the latter half of the twentieth century—today's watches, based on quartz-crystal technology, are accurate to a second a month. And the advent of the atomic clock and space satellites has brought worldwide time synchronization to within 20 billionths of a

second. This has made possible navigation to within 10 metres using satellites, milli-arcsecond studies of celestial radio sources, studies of the Earth's rotation and precise knowledge of the Earth's gravitational field. It also affects fundamental research in areas that measure physical constants and in atomic physics and general relativity.

Time must be based on a phenomenon in nature that has a definite repeatability; this defines the frequency of the 'standard'. This frequency can be measured, but it takes an almost continuous comparison between 'standards' to determine and disseminate a unique timescale by which laboratories can synchronize their clocks to produce the unit of time, the second. Further time must also be considered in the framework of general relativity.

The measurement of time was based originally on astronomical phenomena such as the rotation of the Earth and the motion of the planets. Since 1967, however, it has been based on an atomic-frequency standard — the second is defined by the use of a caesium-133 atom. Our daily lives are governed by civil time, which is based on Coordinated Universal Time (UTC). This is the time at the longitude of the Greenwich meridian, from which the time zones are offset to give local civil time, which follows the rising and setting of the Sun. UTC is International Atomic Time (Temps Atomique International, or TAI) offset by an integral number of seconds in order to keep it in near-synchronization with the rotation of the Earth. As the Earth's

rotation is slowing down by one hour every 1,000 years, 'leap seconds' are added to UTC (at the end of either June or December) when the discrepancy of the Earth's rotation with TAI is 0.9 seconds.

The Measurement of Time gives an excellent, detailed introduction to all aspects of time and frequency. Topics covered include principles, measurements, models, the evolution of time definitions, clock time, atomic standards, atomic time, astronomical time, ultra-precise time and frequency standards, as well as definitions and advanced time applications. All of these topics are well referenced, making this an excellent source for scientists and students in the field of time measurement as well as for historians of physics and astronomy.

Most of the book should also be accessible to a more general audience. Although there are many other books on this subject for the general public, this is perhaps the most authoritative, as it provides the precise definitions used in this field. Such knowledge could provide an interesting topic at cocktail parties when there is a lull in the conversation.

Ken Johnston is at the US Naval Observatory, 3450 Massachusetts Avenue, NW, Washington DC 20392, USA.

More on time

Calendrical Calculations: The Millennium Edition

by Edward M. Reingold & Nachum Dershowitz
Cambridge University Press, £24.95, \$37.95 (pbk)

When politics colours disease

Dying in the City of the Blues: Sickle Cell Anemia and the Politics of Race and Health

by Keith Wailoo

University of North Carolina Press: 2001.

360 pp. \$34.95 (hbk), \$16.95 (pbk)

Ken Fox

The power of black bodies and their afflictions to "tell a moral tale" about changing racial relations in US society is bound to fascinate, infuriate and inspire readers of *Dying in the City of the Blues*. Keith Wailoo's telling historical narrative chronicles the changing meanings — clinical, scientific, political, symbolic — of sickle-cell disease over the course of the twentieth century in America. Fundamentally, this remarkable text on the social construction of the illness speaks of how ancient, bloody, brutal and enduring the facts of racial disparity in health and care really are in the American experience.

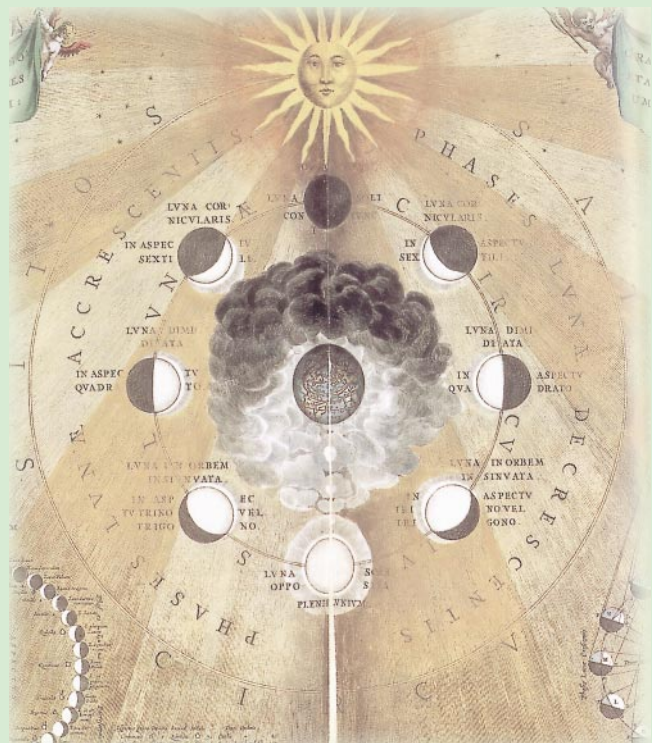
Wailoo's meticulous scholarship excavates a deep and dizzying range of primary documentary resources. Arcane manuscript collections, nearly four dozen technical, scientific and popular journals, newspapers, fiction, blues lyrics, cartoons and sermons are mined to reveal the politics of race and health in America.

What makes this book such an achievement is the author's eye and ear for luminous detail. In one example, Wailoo uses lyrics from bluesman W. C. Handy's 1909 *Memphis Blues* to introduce readers to Edward 'Boss' Crump, mayor and chief architect of city politics in early-twentieth-century Memphis. Crump, "closest thing to a plantation master Memphis had", was "a Democratic Party boss who blended southern paternalism and humor with brutality toward those who disagreed with him". On election day, Crump's henchmen rounded up Handy's audiences, escorted them to the polls, and doled out gifts in exchange for votes. All the while, Crump's regime was viciously policing the boundaries of Jim Crow segregation. Wailoo thus aptly characterizes Memphis blacks as both "pawns and political forces in the urban power struggle". In another example, Wailoo brings Memphis's famous Beale Street to life. Once a potent centre of black capitalism, Beale Street was a major theatre in the holy war waged by "Negro doctors" against "sidewalk barkers" and sellers of "roots, herbs, grape leaves, rabbits' feet, wine and whiskey concoctions".

Through Wailoo's window, the politics of race and health are inextricably bound to class tensions, and central to his interpretive framework is the conception of sickle-cell disease as a 'commodity'. The history of the illness may be traced along its shifting positions

Clock-watching through the ages

Measured time: the phases of the Moon (right), and a pocket sundial from the sixteenth or seventeenth century. From *The Discovery of Time*, edited by Stuart McCready (MQ Publications, £14.99).



within networks of “exchange relationships, where — much like any object — the disease concept and illness experience acquire value and [can] leverage resources, money or social concessions”. As spin and values of the meanings of illness are speculated, negotiated and traded upon, so too do sufferers’ and scholars’ fortunes rise and fall. One of the great advantages of Wailoo’s approach is the way it reveals how social institutions and forces of political economy — more than the triumphant achievements of famous autonomous scholars — shape scientific inquiry and meanings of illness. Hill–Burton grants, in terms of the US Public Health Service Act passed in 1946, have stimulated hospital construction and certain kinds of medical research; philanthropic trends forge particular representations of illness; Medicare and Medicaid alter incentives to serve the sick poor; black power politics shifts cultural representations of black pain; managed care and free-market disciplines within medical markets depreciate the ‘capital’ — fiscal and symbolic — of academic centres. The author shows how social forces find their way “beneath the skin”, into the very marrow of sickle-cell sufferers, transforming options, experiences and outcomes of care.

The text also tells how ideologies come to life. Sometimes those lives — like that of Frankenstein’s monster — are absolutely wretched. For example, the disfiguring power of white racism on medico-scientific discourse is deftly documented by citing a physician, writing in 1932: “The safeguarding of the health of the Negro ...[was] anything but an easy task, for the fight is not against disease, but against physical, mental, and moral inferiority, against ignorance and superstition, against poverty and filth.”

The tale is populated by many public figures of mythic proportions, but no true heroes. Even prominent black social scientists of the age were transfixed by cursed ideologies of white superiority. Wailoo cites sociologist Charles Johnson (1934): “Children die in great numbers and mothers accept their death with a dull and uninquiring fatalism.” Thus, the spectre of the mentally inferior, fatalistic, apathetic and mistrustful black patient whose afflictions are consequences of her flaws — alive even to this day in emerging medical literatures on ‘cultural competence’ — has a lineage as long and shameful as the shadow of the plantation.

The author blows the dust off the many skeletons in the nation’s closet. The ghosts of inequalities past, present and future rattle their chains and point to trouble here, behind and ahead of us. Thankfully, finally, we have, in historian Wailoo, a post-modern Indiana Jones whose rough-and-tumble vitality of practice and piercing gaze allow no place for those old bones to hide. ■

Ken Fox is in the Division of General Pediatrics, Boston University School of Medicine, 91 East Concord, Boston, Massachusetts 02118, USA.

Science in culture

The sum of the parts

Proof, a play by David Auburn

Jonathan Knight

Sophie Germain, daughter of an eighteenth-century Parisian merchant, became fascinated by mathematics at an early age, much to the chagrin of her parents, who thought such concerns were not proper for a young woman. She studied in secret, late into the night, burning a hidden supply of candles. Even after her parents came to accept her strange proclivity, she knew the outside world would not. So when she struck up a correspondence with the great mathematician Carl Gauss, she used a man’s name.

She wrote to him of how she had devised a special kind of prime number and used it to make a start at proving Fermat’s Last Theorem, the famous mathematical teaser that had by then been around for 130 years. Later, when Gauss found out that the author was a woman, he wrote to her that, as a member of the sex that “must encounter infinitely more difficulties than men to familiarise herself with these thorny researches”, she must have extraordinary talents.

Such stories make irresistible fodder for the dramatic artist. In David Auburn’s *Proof*, which began its national tour in the United States in November, the present-day protagonist Catherine foreshadows her own mathematical talent by telling the story of Germain. Catherine is also forced to be reclusive, not because of her sex, but to care for her ageing, schizophrenic father Robert, a mathematician who revolutionised three areas of mathematics before he “went bughouse”.

Although Robert is mentally ill, he is not mad in the fiendish and clinically undefined sense of Dr Frankenstein or Dr Moreau, or even of Seth Brundle in *The Fly*. Robert develops schizophrenia at the age of 25, and the disease halts his career.

Hal, Catherine’s love interest and her father’s former student, now a professor at the University of Chicago, discovers another brilliant mathematical proof locked in Robert’s desk drawer. The question is, who wrote it, Catherine or Robert?

Auburn set out to portray mathematicians and their work as realistically as possible. By consulting real mathematicians, he avoided some of the more hackneyed stereotypes of scientists, and enriched the play with aspects of the culture of mathematics that outsiders might not know. Yet, unlike the plays of Carl Djerassi, *Proof* does not attempt to educate. For Auburn, mathematics is a theatrical tool, and he uses it with great skill.

Proof gives us a look inside the culture of mathematics when Hal tells Catherine it’s a “young man’s game”. Although the general public may know that maths, like physics and the other ‘hard’ sciences, is a field still dominated by men, they may not know that many mathematicians worry that their best years are quickly behind them. “There’s this fear that your



Father and daughter, mathematicians both.

creativity peaks around 23 and it’s all downhill from there,” Hal tells Catherine. He says many of the older guys use amphetamines to try to stay sharp, and he assumes that he will never amount to much as he is already 28.

In a public interview with Auburn at the San Francisco theatre where the tour opened, host Bob Osserman, a mathematician at the Mathematical Sciences Research Institute in Berkeley, said he basically agreed with that portrayal. Although pill-popping is probably not rampant, as the play suggests, mathematicians do worry about losing their edge, he said. And this is unfortunate because, whether justified or not, that fear diminishes the importance of doing good, solid, non-breakthrough mathematics. “The idea that only brilliant research counts is very destructive,” Osserman said.

Where *Proof* goes beyond the realm of reality is in asking us to believe in Catherine’s amazing mathematical insight. It’s true that in maths, but rarely in other scientific disciplines, great strides can be solo acts. But not even Germain achieved what Catherine says she achieves, a brilliant mathematical proof written in a few months of late nights with no more background than a genius father and a semester of college.

Nevertheless, Auburn uses mathematics as an effective dramatic tool that leaves the audience on the edge of its seat, hoping desperately that Catherine wrote the proof. The truth is not revealed until the final scene. ■
Jonathan Knight writes for *Nature from San Francisco*.

Proof is playing at the Boston Wilbur Theatre until 17 February. The tour reaches the Denver Auditorium Theatre on 18 June.

CHRIS BENNION

Fact and fantasy

The zoology created by our imagination is far outstripped by that of reality.

Sandra Knapp

"I have come to these conclusions by personally leading my pupils on wanderings through the tangled web of nature, in order that I can spur others on to an examination and explanation of nature rather than the reiteration of perceived ideas ... I shall take exception to the tales of actors and the barkings of dogs with equal measure." With this dismissive final paragraph, Carolus Linnaeus notified the readers of the sixth edition (1748) of *Systema naturae*, his compendium of life on Earth, that he would no longer be including imaginary beings in his system.

In all the editions of the *Systema* (the first was published in 1735), Linnaeus specified three kingdoms of nature — "stones, which grow", "plants, which grow and live", and "animals, which grow, live and feel". In the earlier editions, however, he also included, floating in limbo and belonging to none of these kingdoms, the *Paradoxa* — a group including such creatures as the hydra, the fabulous monster known as the manticore and the *Automa Mortis*, the death-watch beetle.

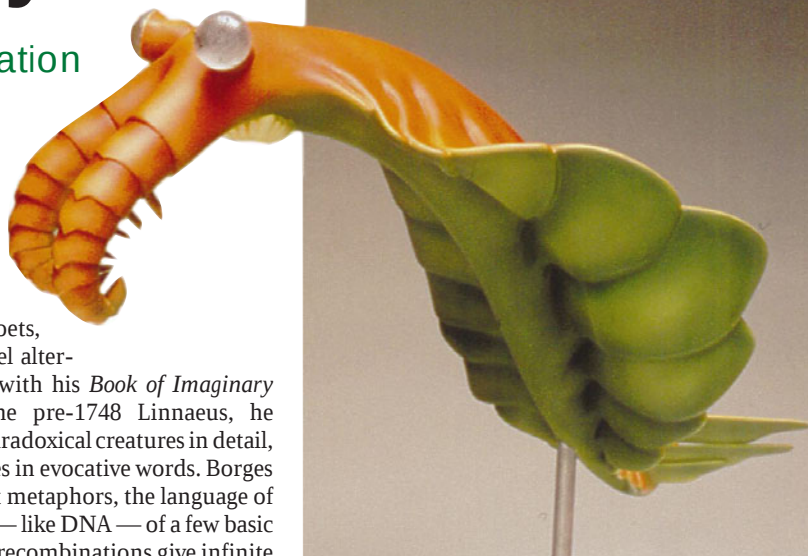
Linnaeus described each member of the *Paradoxa* with great care, as if unsure that they did exist. By 1748, however, he had clearly decided that the *Paradoxa* had to go. They were actors' tales and as meaningless as the barkings of dogs, mere noise in the system, with no place in his carefully observed and documented scheme of life on Earth. He reassigned some of the names used for the *Paradoxa* to living organisms — 'Pelecanus' became a bird, 'Hydra' a coelenterate and 'Draco' a reptile. Others he made synonymous with real genera — 'Satyrus' (the satyr), for example, was considered a monkey and *Rana-Piscis* (tadpoles) a type of salamander. Some of the *Paradoxa* just disappeared without trace, such as the 'Boramez', or 'vegetable lamb'.

In describing the diversity of nature, scientists create new words for that which exists but is outside our imagination.

Jorge Luis Borges, one of the most learned and lyrical of twentieth-century authors and poets, created a parallel alternative zoology with his *Book of Imaginary Beings*. Like the pre-1748 Linnaeus, he described the paradoxical creatures in detail, painting pictures in evocative words. Borges maintained that metaphors, the language of poems, consist — like DNA — of a few basic patterns whose recombinations give infinite diversity. Every author's reinvention of the metaphor allows readers to see and feel new things. The imaginary beings of Borges, although sharing a name with those of Linnaeus, free the reader to explore the diversity of the human imagination. Borges, in his introduction, wrote that "anyone looking at the pages of the present handbook will soon find out that the zoology of dreams is far poorer than the zoology of the Maker". At first this seems paradoxical in itself; how can we not imagine a world of beings that is infinite, by "evolv[ing] an endless variety of monsters — combinations of fishes, birds, and reptiles, limited only by our boredom or disgust?" Surely our imagination can devise a world more strange than the one we inhabit.

But Borges was certainly right — the zoology of the imagination is far poorer than that of reality. Who could imagine a plant that flowers under water (*Thalassia*, a sea grass), a phylum of creatures that live only on the 'lips' of lobsters (see *Nature* 378, 711–714; 1995), or *Anomalocaris* and *Hallucigenia*, the seemingly impossible Precambrian creatures of the Burgess Shale? Evolution and natural selection, the forces that shape diversity in nature, are far more potent engines for innovation than even the most fertile human imagination.

Our imagination is limited perhaps by the nature of language — by words themselves. In a series of lectures Borges gave at Harvard University in the late 1960s and which have recently been edited from long-lost tape recordings (*This Craft of Verse*), he contends that words themselves are not one-for-one symbols for abstract thoughts; they are, in fact, deeply rooted in the concrete, and they evolve. A not surprising notion, perhaps; but, taken in conjunction with his idea that "a nation



Strange but true: a model of *Anomalocaris*, based on fossils from the Burgess Shale.

evolves the words it needs ... language comes from the fields, the sea, from rivers, from night, from the dawn", it reveals how science and poetry, often thought to be uneasy bedfellows, have a similar basis in empirical observation.

The vocabulary of science is also firmly rooted in the concrete — in observation and careful documentation of structure and events. Linnaeus excluded the *Paradoxa* from his kingdoms of nature because he had no concrete evidence of their existence. These 'imaginary beings' are metaphors — reinventions and recombinations of existing organisms. The 'real beings' of nature must be described to even begin to be understood. Both poets and scientists choose words with precision. But in describing the diversity of nature, scientists create new words — a new vocabulary of meanings for concrete objects — for that which exists but is outside our imagination. In so doing, they provide the words with which poets fashion dreams.

Sandra Knapp is in the Department of Botany, Natural History Museum, Cromwell Road, London SW7 5BD, UK.

FURTHER READING

Borges, J. L. with Guerrero, M. (transl. di Giovanni, N. T.) *The Book of Imaginary Beings* (E. P. Dutton, New York, 1969). [First published as *Manual de zoología fantástica* in Mexico, 1957; *El libro de los seres imaginarios* in Buenos Aires, 1967.]
Borges, J. L. (ed. Mihailescu, C.-A.) *This Craft of Verse* (Harvard Univ. Press, 2000).
Briggs, D. E. G., Erwin, D. H. & Collier, F. H. *Fossils of the Burgess Shale* (Smithsonian Inst. Press, 1994).
Linnaeus, C. *Systema naturae* (Univ. Uppsala; 1st edn, 1735; 6th edn, 1748).

NATURAL HISTORY MUSEUM, LONDON

Mind-grasping gravity

Victor Smetacek

Imagine yourself standing at the edge of a precipice, looking down at its foot, and then crouching at the same place on all fours. The difference between the two sensations is the difference between being human and being a quadruped. Clearly, latent anxiety is inherent in the precarious, human mode of bipedalism, which balances a vertical vertebral column on straight legs, with no tail for support.

Balance is so central to every activity, both of the body and the mind, that it is simply taken for granted. It is imbalance (disturbance, perturbation) that captures attention, be it fear of falling, the mental struggle to balance an equation, or the moral urge to right an injustice. As the concept of balance applies smoothly across the entire range of human endeavour, it would be parsimonious to assume a direct connection between the concrete and the abstract, on the basis of compatible neural hardware in the brain. Is the mind's balance, and hence its functioning, derived from that of the body?

Aristotle did not list balance as one of the explicit senses, although it is based on sensory organs. In contrast, Eastern philosophy is explicitly balance-based. Balance is to gravity as vision is to light, or hearing to sound; but whereas light and sound fields vary, Earth's gravitational field is constant.

Because sense organs

senses the gravitational field with great precision when the body moves.

Three independent organ systems enable the body to maintain balance. To experience how they interact with and compensate for each other, stand close to a wall, on one leg, with your arms dangling, and then shut your eyes; repeat the experiment but touch the wall lightly with a fingertip beforehand. Clearly, we rely on vestibular, visual and somatosensory systems to get our bearings in relation to gravity.

The vestibular organs of the inner ear sense gravity directly, but also as a deviation from the vertical and as self-motion. Balance and momentum signalled by this complex system are manifested in the body's centre of mass — the lower gut — as experienced on heaving ships and rollercoasters. These sensations can also be evoked, as in nightmares.

The eyes also sense and appreciate balance and mass. We enjoy watching dancers, athletes and acrobats (but also clowns), and the mass and symmetry of monumental buildings (or leaning towers) fill us with awe (or other emotions). Beauty, symmetry and balance evidently go hand in hand — there is more to the eye of the beholder than just vision.

The somatosensory system comprises a variety of receptors in the skin, muscles and skeleton which sense gravity as pressure and weight. Body awareness (proprioception) is part of this system. Although the arms are decoupled from locomotion, hand-held tools such as a cane (equivalent to touching the wall) or an acrobat's balancing rod significantly enhance the body's ability to balance.

Each sensory system provides independent, but integrated, coordinates for the body (including the hands) to orientate itself. Research on the vestibular cortex is in its infancy, but its multiple representation, its intimate interaction with visual and sensorimotor cortices and its right-hemispheric dominance distinguish it from other sensory systems. Recent studies indicate that the vestibular system is involved in self-perception and cognition. The human cerebellum, a central organ of balance and also of fine motor skills, contains five times as many neurons as the cerebrum but has received much less attention. Additional functions are only now coming to light.

As balance is central to every directed movement, evolving fine motor skills is synonymous with fine-tuning the sense of balance. Human evolution can be characterized as stages in differentiation and refinement of our balancing abilities. Our lineage first learned to balance bodies on feet, then tools in hands, and most recently,

Balance

Is the mind's balance, and hence its functioning, derived from that of the body?

instruments and aircraft with eyes.

Refining balancing ability improved tool production and use, but also resulted in a form of perception that is linked to the hands and decoupled from the body. Whereas whole-body proprioception and personal viewpoint (the body's sense of balance) is subjective and private, the hands weighing different objects as the pans of a balance (say one big stone in the left, and two small ones in the right) create a demonstrable, verifiable quality (the balance between objects) that can be judged externally and objectively. With this 'disembodied' sense of balance, the principles of constancy and equivalence (the basis of common-sense logic) could be grasped, understood and communicated, in successive stages of evolution. Eventually, measuring rods, pendulums, levers and balances, which are mechanical projections of the body, could be transformed into abstract projections within the mind, unified by an innate understanding of gravity.

Just as there is a mind's eye and a mind's ear, there must also be a mind's gravity, based on each sensory system either on its own or in concert. This is the mind's space-time coordinating system, in which mass, balance and momentum — the substrates of science — are sensed. Archimedes, Newton and Einstein, among a host of others, have shown that there is more to insight than just vision or words.

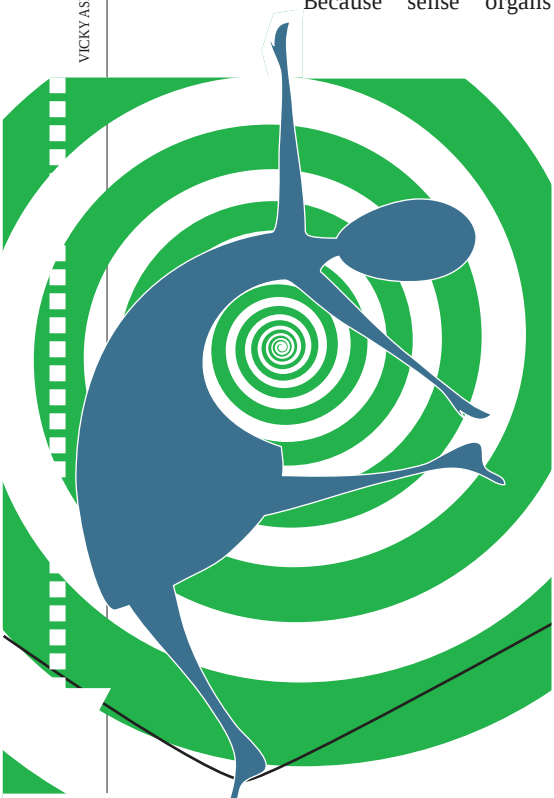
The bipedal apes striding across the savannah with head held high evolved a very different proprioception and world view to their slouching cousins. Our ancestors dared to challenge gravity by standing up to it and we continue to do so, with our bodies and tools, minds and machines, balancing our way onwards and upwards, both literally and figuratively.

Victor Smetacek is at the Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany.

FURTHER READING

Brandt, T. & Dieterich, M. *Ann. NY Acad. Sci.* **871**, 293–312 (1999).
Heck, D. & Sultan, F. *Spektrum der Wissenschaft* October 2001, 36–44 (2001).
Lackner, J. R. *et al. Exp. Brain Res.* **126**, 459–466 (1999).
Ramachandran, V. S. & Blakeslee, S. *Phantoms in the Brain: Probing the Mysteries of the Human Mind* (Quill, William Morrow, New York, 1999).

VICKY ASKEW





The investment forecast

Reiner Schnur

New studies predict that the risk of extreme rainfall over Europe and Asian monsoon regions is increasing, with more floods likely worldwide. Such long-range forecasting is pushing at the limit of current climate models.

The world's climate changed during the twentieth century and is continuing to do so. At least part of this change is caused by anthropogenic emissions of greenhouse gases and sulphate aerosols, according to the latest report of the Intergovernmental Panel on Climate Change (IPCC)¹. Changes in extreme climate², such as hot spells, droughts or floods, potentially have a much greater impact on society than changes in mean climate, such as summertime temperature averaged over several decades^{3,4}. So the ability to assess future risks associated with extreme events is increasingly important to policy-makers.

Two new studies (on pages 512 and 514 of this issue)^{5,6} report evidence that increases our confidence in observed and projected changes in extreme rainfall and flooding. Notably, Palmer and Räisänen⁵ link the results of several climate models with a decision-making tool, and show that some approaches to forecasting are more valuable than others. Such tools will become increasingly important to both climate-modellers and policy-makers. For their part, Milly *et al.*⁶ show that the frequency of severe floods in large river basins has increased during the twentieth century (Fig. 1), with only a small likelihood that these changes are due to natural climate variability.

Palmer and Räisänen⁵ analyse the output of 19 climate models and estimate that, over the next century, very wet winters will be up to five times more likely than today for much

of central and northern Europe. They also predict that the probability of very wet summers in the Asian monsoon region will rise by a similar magnitude, increasing the risk of flooding. By applying a simple 'cost-loss' analysis to a hypothetical investment decision, they show that for extreme events it is better to consider the changing frequency of extreme events in a set of different climate-change projections (the 'probabilistic' approach), than to rely on a single projection or on the averaged projection from several climate models (the 'consensus' approach used regularly by the IPCC¹).

To estimate future climate change caused by anthropogenic emissions of, say, CO₂, climate scientists use numerical models that simulate the effects of atmospheric changes on climate. But such climate models are inherently imperfect owing to physical processes that are either not completely understood or have yet to be adequately represented because of limited computational power. So there is a fundamental modelling uncertainty attached to a given simulation from any climate model, over and above any naturally occurring climate variability. The consensus approach to solving this problem is to assume that the influence of modelling uncertainty can be limited by averaging several different models that differ slightly in their representation of climate. The IPCC report¹ relies heavily on this approach to provide 'best-guess' projections of future climate change that describe the

Figure 1 (above) **Heavy floods, such as this one in Oregon City, United States, in 1996, can have a huge impact on human affairs. Milly *et al.*⁶ show that the frequency of severe floods has increased during the twentieth century, with only a small chance that these changes are caused by natural climate variability — anthropogenic climate change seems more likely.**

features common to an 'ensemble' of different model simulations. But there is a problem if this consensus approach is applied to the analysis of extreme climate events (Fig. 2, overleaf). Using a consensus projection for planning purposes would inevitably underestimate the risk associated with extreme rainfall.

Palmer and Räisänen take a more explicit approach to the problem of accounting for modelling uncertainty and adopt probabilistic techniques that are now routinely used in short- and medium-range weather forecasting, in which uncertainties owing to incomplete knowledge of initial conditions have to be considered⁷. The authors analyse 80-year projections from 19 different global-climate models that include interactions between the oceans and the atmosphere. First, the models were run at a fixed CO₂ concentration to define a 'baseline' ensemble of simulations corresponding to today's climate conditions. Next, they were run with a 1% per year increase in CO₂, which is slightly faster than is predicted for the twenty-first century, but seems reasonable because

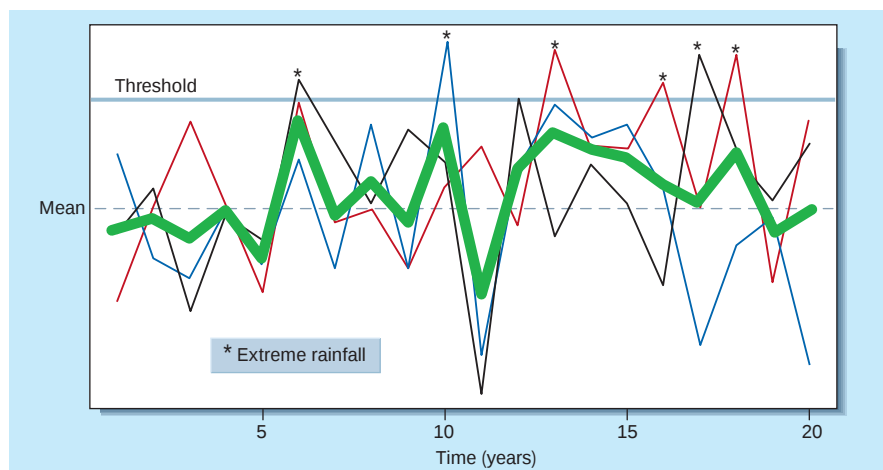


Figure 2 The pitfalls of forecasting extreme climate events. Palmer and Räisänen⁵ show that it is better to consider the frequency distribution of such events among a number of different climate simulations than to use only a single simulation or to take an ensemble (consensus) average of several models. For example, the probability of summertime rainfall at a given location exceeding a certain threshold (solid horizontal line) will be unrealistically low in an ensemble average (green line) of several simulations (thin lines) because of the smoothing effect of averaging. In the example above, this probability would be zero for the ensemble average, but positive in each one of the three simulations. Using only a single simulation for forecasting the probability of extreme rainfall would give answers of between 5% and 15% in this example (corresponding to one and three flooding events over 20 years), depending on the model used.

these model integrations neglect the contributions of other greenhouse gases.

By comparing the relative frequencies of extreme events in the baseline and climate-change ensembles, Palmer and Räisänen obtain a measure of the changing risk associated with anthropogenic climate change. Although this technique is not new, this is the first time it has been applied to a large multi-model ensemble in a climate-change context to assess the frequency of extreme events. This approach is better at accounting for model and sample uncertainty, although systematic errors common to all models cannot be avoided. To date, most studies have used single projections or ensembles from one climate model (see, for example, Milly *et al.*⁶ or Kharin and Zwiers⁸).

This work⁵ also demonstrates the economic value of probabilistic forecasts of extreme events to policy-makers. By applying a simple cost-loss decision model to a hypothetical long-term investment decision in a region prone to flooding, the authors show that it is consistently better to base decisions on a probabilistic forecast than on a single simulation. (For assessing extreme risks, a consensus forecast is never better than simply assuming that the frequency of events will remain constant, because ensemble averaging always underestimates extreme events; Fig. 2). Although the cost-loss model used by Palmer and Räisänen is too simple to be used in real-world decision-making, their approach suggests a reliable way to make better use of ensemble predictions in climate-change research and risk assessment.

Studies such as these make a strong case for improving computational resources in

climate research. Much larger ensembles are needed for more reliable estimates of extreme events, particularly for very rare events that have a great potential for a disastrous impact on a country's economy and inhabitants. Multi-model ensembles of the size used by Palmer and Räisänen are available only for future changes in CO₂ concentration. Similar model integrations will be needed for a more complete picture of the factors contributing to climate change,

such as other greenhouse gases and sulphate aerosols, including any uncertainties attached to these agents.

Today's climate models are not good at predicting extreme climate events in local areas, such as flooding in a given river basin, because they are limited in their resolution to a coarse grid size of about 200 kilometres. In Palmer and Räisänen's study, only a few river basins — such as those of the Brahmaputra and Ganges in India and Bangladesh — are large enough to allow their results for extreme seasonal precipitation to be translated into a heightened risk of river flooding. This is also why Milly *et al.*⁶ use only the largest basins in their analysis of great floods. For the average river basin, climate-change simulations would need a much higher resolution of tens of kilometres, but this will not be available for quite some time. Until computational power increases significantly, climate scientists will have to patch models together, taking the results of ensemble climate projections, for example, and inserting their output into a high-resolution hydrological model for a specific river basin. ■

Reiner Schnur is at the Max Planck Institute for Meteorology, Bundesstrasse 55, 20146 Hamburg, Germany.

e-mail: schnur@dkrz.de

1. Houghton, J. T. *et al.* (eds) *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, 2001).
2. Meehl, G. A. *et al.* *Bull. Am. Meteorol. Soc.* **81**, 413–416 (2000).
3. Easterling, D. R. *et al.* *Science* **289**, 2068–2074 (2000).
4. McCarthy, J. J. *et al.* (eds) *Climate Change 2001: Impacts, Adaptation and Vulnerability* (Cambridge Univ. Press, 2001).
5. Palmer, T. N. & Räisänen, J. *Nature* **415**, 512–514 (2002).
6. Milly, P. C. D., Wetherald, R. T., Dunne, K. A. & Delworth, T. L. *Nature* **415**, 514–517 (2002).
7. Palmer, T. N. *Prog. Rep. Phys.* **63**, 71–117 (2000).
8. Kharin, V. V. & Zwiers, F. W. *J. Clim.* **13**, 3760–3788 (2000).

Cancer

The molecular outlook

Carlos Caldas and Samuel A. J. Aparicio

Many breast-cancer patients receive unnecessary treatment for possible tumour spread after the removal of a primary tumour. Molecular profiling should offer more accurate predictions of who needs such treatment.

The first step in rationally treating a disease is to assess the patient against a classification of diseases, the results being used to predict the person's response to various therapies. The effectiveness of the process depends on the quality of the classification. Cancer is no exception to this rule, and the advent of microarray methods to analyse DNA, RNA or proteins from tumour cells has started to refine the classification of cancer¹ to levels that classical methods have been unable to reach.

The report by Friend and colleagues on page 530 of this issue² provides a good example of how such 'molecular profiling' is improving cancer classification. It shows

that the results of gene-expression profiling of breast tumours, carried out after they have been surgically removed, can be used to predict which patients will develop clinical metastases (the spread of the tumour to other sites, where secondary tumours then develop). If molecular forecasting of the outcome of cancer is indeed possible, as this work suggests, it is a significant advance on existing prognostic methods.

Breast cancer is common: in the United States and Britain, one in ten women will develop the disease, and half of those will die with it. Treatment for individual patients is chosen according to various criteria, such as the extent of tumour spread (also called

staging, which involves determining tumour size, whether cancer cells have spread to the axillary lymph nodes and how many nodes are involved, and whether distant clinical metastases are present). Other factors are the patient's age, the degree to which the cancer cells resemble their normal counterparts, the 'tumour grade', and whether the cancer cells express the oestrogen receptor. In women with no evidence of metastasis, the mainstay of treatment aimed at curing the disease is removal of the breast tumour (and often also the axillary lymph nodes) and radiotherapy. Unfortunately some of these patients later develop clinical metastases. There is a need, then, to identify women who, after surgery, will require further ('adjuvant') therapy for the microscopic deposits of cancer cells that may have already spread from the primary tumour^{3,4}.

Adjuvant therapy uses pharmaceutical agents — such as oestrogen modulators or cytotoxic drugs — that reach cancer cells through the bloodstream. So it is not surprising that these treatments frequently have toxic side effects. At the moment, identifying women who might need such treatment relies on various clinical and histopathological indicators, as described above. Even taken together, however, these indicators are only

poorly predictive. So, to save a sizeable but small percentage of lives, many patients who would have been cured by surgery and radiotherapy alone go on to receive unnecessary and frequently toxic treatment.

Friend and colleagues² used gene-expression microarrays to analyse some 25,000 genes in 78 primary breast cancers from young women. Microarrays are glass slides or chips, carrying DNA representing thousands of genes, to which labelled RNA isolated from cancer cells is applied to determine gene activity. These patients were selected to be 'node negative' at the time of diagnosis — that is, they did not have axillary lymph-node metastases — and, with the exception of five individuals, they had not received adjuvant therapy. The gene-expression data were analysed using statistical methods, with the aim of identifying gene clusters that correlate with the patients' known clinical outcome.

The most interesting results came from a three-step 'supervised' clustering method for tumour classification. The supervised method uses existing information, in this case knowledge of clinical outcomes, to guide the identification of gene clusters. The authors identified an optimal number of 70 marker genes — the 'prognosis classifier' — which correctly predicted the later appearance or absence of clinical metastasis in 65 of the 78 patients; this 'prediction' was, of course, retrospective, because the outcomes of the group studied were already known. The authors then determined an 'optimized threshold' that would result in no more than 10% of 'poor prognosis' patients (that is, patients who had clinical distant metastasis within five years of surgery) being misclassified, which produced only a small drop in overall predictor accuracy.

Friend and colleagues then applied this optimized threshold to a similar but independent set of 19 breast-cancer patients, which resulted in only two incorrect classifications. The predictive power of the 'poor prognosis' gene-expression signature for clinical metastasis was much greater than that of currently used approaches.

How would these results look if they apply more widely? Analysis of Friend and colleagues' data provides the necessary information. Using the current predictive criteria^{3,4}, 70–91% of the patients cured with surgery and radiotherapy would be unnecessarily advised to receive adjuvant therapy. In contrast, using molecular forecasting, only a quarter of those patients would have been advised to undergo such treatment. Importantly, the number of patients with poor prognosis, wrongly assigned to the 'no treatment' group, would not differ significantly using conventional or molecular criteria. Because most node-negative patients are cured by surgery and radiotherapy alone, the effect of better selection is potentially staggering, as indicated by the red bars in Fig. 1.

These findings are tremendously promising. But history tells us to be cautious, and the widely accepted guidelines for evaluating breast cancer remain essential⁵. Cancer classification based solely on gene-expression profiling may well be feasible^{6–9}, but this study² was only a pilot and had a small sample size. Furthermore, a retrospective tumour bank was used, and only tumours where cancer cells constituted more than 50% of the overall cell composition were analysed, so there was potential for biological bias. Finally, of course, the method and the reproducibility of the results need to be tested in the clinic.

What about future priorities? First, it has become a matter of urgency to judge the success of different kinds of treatment against tumour categories at the DNA, RNA or protein level. It is increasingly obvious¹⁰ that classical histopathological and clinical criteria are deficient, and that molecular classifications could dramatically improve patient care. But for this to happen, tumour material and clinical data will have to be collected carefully and efficiently, which has both ethical and cost implications.

Second, clinical trials will need to include tissue collection as a key component, thereby allowing future treatments to be evaluated by molecular techniques. Studies that do not include these analyses may mean that crucial new insights go begging. Third, transfer to the real world of patient treatment will require the development of straightforward techniques for use in the clinical laboratory. Ideally, these methods will involve only a limited selection of predictive classifiers, and so be suited to high-throughput automated analysis. Obviously, these are priorities not just for breast cancer but for other types of malignancy.

During the twentieth century the selective use of antimicrobial agents, following pathogen isolation and sensitivity testing, transformed the treatment of infectious disease. If findings such as those by Friend and colleagues² can be generalized, we can likewise hope that cancer treatment will be vastly improved by better predicting the response of individual tumours to therapy.

Carlos Caldas and Samuel A. J. Aparicio are in the Cancer Genomics Program, Department of Oncology, University of Cambridge, Hutchison/MRC Research Centre, Addenbrooke's Hospital, Cambridge CB2 2XZ, UK.
e-mail: cc234@cam.ac.uk

1. Golub, T. R. *et al.* *Science* **286**, 531–537 (1999).
2. van't Veer, L. J. *et al.* *Nature* **415**, 530–536 (2002).
3. Goldhirsch, A., Glick, J. H., Gelber, R. D. & Senn, H.-J. *J. Natl Cancer Inst.* **90**, 1601–1608 (1998).
4. National Institutes of Health Consensus Development Panel. *J. Natl Cancer Inst.* **93**, 979–989 (2001).
5. McGuire, W. L. *J. Natl Cancer Inst.* **83**, 154–155 (1991).
6. Alizadeh, A. A. *et al.* *Nature* **403**, 503–511 (2000).
7. Bittner, M. *et al.* *Nature* **406**, 536–540 (2000).
8. Perou, C. M. *et al.* *Nature* **406**, 747–752 (2000).
9. Khan, J. *et al.* *Nature Med.* **7**, 673–679 (2001).
10. Aparicio, S. A., Caldas, C. & Ponder, B. *Genome Biol.* **1**, 1021.1–1021.3 (2000).

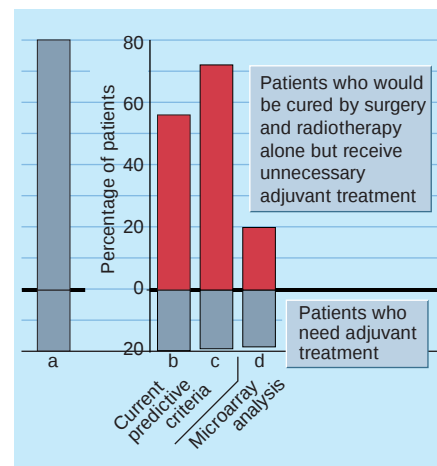


Figure 1 Molecular forecasting and cancer therapy. a, Of young women diagnosed with breast cancer but whose lymph nodes are unaffected, 80% will be free of disease for five years following surgery and radiotherapy alone. The remaining 20% will develop metastases, and would benefit from receiving further (adjuvant) therapy. The problem is identifying the patients who fit into each group. b, c, Proportion of patients who, on two types of current predictive criteria (b, NIH Consensus, and c, St Gallen)^{3,4}, would be advised to receive further therapy. d, The situation if the molecular predictor described by Friend and colleagues² were validated and used for the basis of advice. The red bars show the proportion of patients who would be cured by surgery and radiotherapy but would then receive unnecessary (and frequently toxic) adjuvant treatment.

Chemistry

One generation at a time

Christopher Gorman

Controlled binding of atoms or molecules within a larger structure could offer new routes to drug delivery or nanoscale materials. Synthetic dendrimers can be tailored to bind ions in just such a regulated manner.

The long chains of molecules that make up a polymer tend to be arranged in a linear way, often resembling lengthy, tangled strips of molecular spaghetti. But if certain short chains are brought together in a highly controlled way, a wholly different, organized structure can result. Such artificial molecules are called dendrimers, named for the tree-like branched structure of the polymer chains that radiate from the molecule's central core. Starting with a core, branched monomer molecules are added sequentially to form concentric layers, or 'generations' (Fig. 1). By manipulating the molecular structure generation by generation, chemists can tailor the properties of the overall dendrimer.

One of the best examples of tailoring so far is the creation of an electric-potential gradient between a dendrimer's inner and outer generations¹. Such molecules can 'harvest light' — photons absorbed by the outer generations over a broad spectral range are channelled to the molecule's core and re-emitted. On page 509 of this issue, Yamamoto *et al.*² describe a different kind of gradient — one that influences the way in which guest ions bind to a dendrimer's generations. Understanding and controlling such processes could have implications for the design and assembly of new nanomaterials.

Yamamoto *et al.* examined a dendrimer built from four generations of imines — monomers that contain nitrogen — and assessed the binding, or 'complexation', of tin (Sn^{2+}) ions to the nitrogen atoms within the structure. Intuitively, one would expect the ions to complex with the first imines that they come across — but instead they head for the imines closest to the dendrimer's core. The complexation process then continues, generation by generation, until the outermost layer is reached. Yamamoto *et al.* suspect that a gradient of basicity — the willingness of a chemical group to accept protons, or positive charge — between the inner and outer layers may be behind this unusual behaviour.

The stepwise nature of the complexation process was uncovered by monitoring the dendrimer's absorption of radiation at ultraviolet and visible wavelengths as increasing amounts of Sn^{2+} ions were added to the solution. As a reaction typically proceeds, the absorption spectrum of the mixture evolves as the reactant disappears and the product forms. But there is usually a point at which the magnitude of the optical absorp-

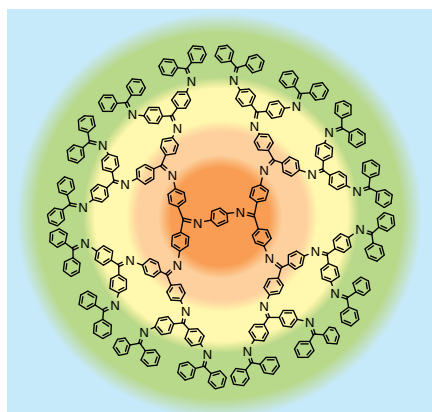


Figure 1 The hyperbranching architecture of the organic molecules known as dendrimers forms a unique structure of topologically concentric layers, or 'generations', around the molecular core. Yamamoto *et al.*² show how tin (Sn^{2+}) ions bind to the nitrogen atoms in this four-generation dendrimer in a stepwise manner, starting at the innermost generation. They note that this complexation process can be controlled by adjusting the relative basicity of the generations.

tion of the reactant (here, that of the uncomplexed dendrimer) and that of the product (the first-generation-complexed dendrimer) are the same. Called the isosbestic point, this is the point at which the level of absorption remains unchanged as the reaction proceeds. Although the spectra of two distinguishable species might intersect once to give an isosbestic point, it is extremely unlikely that the spectra of three or more distinguishable species will mutually coincide. So the stability of the isosbestic point is good evidence that a reaction is proceeding without forming intermediate or multiple products.

Yamamoto *et al.*² saw four successive isosbestic points. The first occurred when the equivalent of two Sn^{2+} ions per dendrimer molecule was added; the second when the next four equivalents were added. The remaining points corresponded to the addition of 8, then 16 more equivalents. This numerical pattern matches the number of imine sites in successive generations of the molecule, working from inside to out, and shows that complexation occurs in a stepwise manner.

In the second stage of the experiment, the researchers added an electron-withdrawing chemical group to the dendrimer's core, which distorted the relative basicity of the imines at each generation. Tracing the

isosbestic points as Sn^{2+} ions were added, Yamamoto *et al.* saw a subtle change in the complexation process — the imine sites in the second generation were complexed first, followed by those in the third and fourth generations. The innermost generation, now with lowered basicity caused by the electron-withdrawing group, complexed last of all.

Something unique to the dendrimer causes this stepwise complexation. Yamamoto *et al.*² showed that it does not occur in a dendron (a single branch of the dendrimer) or in a linear oligomer. Why the dendrimer architecture should have this effect is not clear. But the observation of the stepwise process — and of a means to control it — is intriguing and potentially useful. Although self-assembly is a favoured strategy for building organized molecular structure and function at the nanometre-scale, it is not yet a general synthetic method. The stepwise complexation seen by Yamamoto *et al.* is essentially a form of directed self-assembly and could be a powerful method to control nanometre-scale structure. For example, the fabrication of nanoparticles within dendrimers³ might be extended to the preparation of new types of core-shell structure⁴⁻⁶ — or perhaps new types of multilayered nanoparticle structure.

One especially active area is the investigation of dendrimers for use in molecular recognition⁷. This uses large molecules as hosts to recognize and manipulate smaller, guest molecules to catalyse a reaction, for example, or to transport the guest molecules into new environments. In drug delivery⁸ and medical imaging⁹, the type of complexation studied by Yamamoto *et al.* could open up new opportunities. Dendrimers that bind to antibodies in their outer regions and to therapeutics in their inner regions, for instance, might offer an improved way to target cancerous cells with poisonous antitumour drugs.

There is still much to be learned about dendrimers. What conformations do they adopt and how does this influence the way in which the generations are laid out in space?¹⁰ We will have to answer such questions in order to exploit these molecules as successful nanostructures or nanoscale delivery agents. But if nothing else, Yamamoto *et al.* have illustrated an intriguing new phenomenon. Their studies suggest that the construction of dendrimers is a promising way of building function into macromolecules. ■

Christopher Gorman is in the Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, USA.

e-mail: chris_gorman@ncsu.edu

1. Hecht, S. & Fréchet, J. M. J. *Angew. Chem. Int. Edn* **40**, 74–91 (2001).
2. Yamamoto, K., Higuchi, M., Shiki, S., Tsuruta, M. & Chiba, H. *Nature* **415**, 509–511 (2002).
3. Zhao, M. & Crooks, R. M. *Angew. Chem. Int. Edn* **38**, 364–365 (1999).
4. Yoffe, A. D. *Adv. Phys.* **50**, 1–208 (2001).
5. Thurmond, K. B. II, Kowalewski, T. & Wooley, K. L. *J. Am. Chem. Soc.* **119**, 6656–6665 (1997).



100 YEARS AGO

In a paper on the energy of the universe in the *Revue scientifique*, M. I. Skvortzow discusses the influence of electrical phenomena in cosmogony. He considers that in the past history of the earth, and of other celestial bodies, electrical and chemical energy have originally played the most important part, and that heat energy has become more and more important in proportion as the earth has assumed a more material form... The heat of the earth M. Skvortzow attributes to electric currents circulating mostly near the surface; the interior of the earth, on the other hand, he thinks may be as cold as the greatest depths of the ocean. Changes in the aspect of the earth, as well as meteorological phenomena, are attributed to electric currents induced by solar influence. The temperatures of different planets are considered to depend less on their distance from the Sun than on their reserve of energy and on the currents which the Sun induces in them in virtue of their axial and orbital motions. Will this theory of the electromagnetic origin of the earth's heat reconcile the two opposing views on the age of the earth?
From *Nature* 30 January 1902.

50 YEARS AGO

A new periodical is apt to arouse suspicion and evoke the question, "Is it really necessary?" It may be said that two conditions must be fulfilled for a newcomer to be regarded as a journal of importance: the first is that the contents must be of high quality; and the second that it serves as a medium for publishing work which might not otherwise reach those to whom it is of interest. As regards the first requirement, there can be little doubt of its fulfilment by *Chemical Engineering Science: Genie Chimique*, an international monthly the first number of which appeared last October... There are seven editors, all chemical engineers of high standing, residing in France, Great Britain, Italy, Switzerland, Belgium, Holland and Norway, and an advisory board of nineteen from eleven countries. Six original papers of high quality occupy fifty quarto pages very clearly printed by the Würzburg University Press, and it is hoped to publish in English, French or German not only papers dealing with the principles of chemical engineering but also detailed notes on advances in industrial processes.
From *Nature* 2 February 1952.

6. Marinakos, S. M. *et al.* *J. Am. Chem. Soc.* **121**, 8518–8522 (1999).
7. Zeng, F. & Zimmerman, S. C. *Chem. Rev.* **97**, 1681–1712 (1997).
8. Kukowska Latallio, J. F. *et al.* *Proc. Natl Acad. Sci. USA* **93**,

- 4897–4902 (1996).
9. Liu, M. J. & Fréchet, J. M. J. *Pharm. Sci. Technol. Today* **2**, 393–401 (1999).
10. Gorman, C. B. & Smith, J. C. *Acc. Chem. Res.* **34**, 60–71 (2001).

Bacterial behaviour

Function by serendipity

Dagmar Ringe

Crystallography has provided an unusual route to the characterization of a bacterial signalling agent: the molecule concerned has been accidentally caught in the clutches of its receptor.

In most of biology, research projects usually follow a hypothesis-driven path that starts with the identification of some activity or other and leads to knowledge of the detailed functions of the molecules responsible. Ultimately, complete characterization of function requires information at many levels of organization, and experiments of many kinds.

For instance, in the case of the signalling pathway described by Chen *et al.* on page 545 of this issue, the activity that was originally identified is the ability of bacteria to communicate with each other. A complete description of the phenomenon would include the chemical composition of the signalling molecule or molecules, the atomic structures of the target receptor(s) on the bacterial cell wall, how the molecule interacts with the receptor, and a description of the ensuing pathways inside the cell that produce the response.

In their paper, the authors describe two parts of this characterization: the chemical composition of the signalling molecule and the atomic structure of the target receptor (Chen, X. *et al.* *Nature* **415**, 545–549; 2002). The serendipitous result of their efforts to obtain one led to the characterization of the other. Determining the three-dimensional structure of the sensor protein, LuxP, accidentally provided the structure of the signalling molecule, AI-2, because it co-purified with the sensor. This enabled Chen *et al.* to identify AI-2 as a boron-containing molecule — a furanosyl borate diester. This remarkable observation suggests a role for boron in the cell, as well as providing the identity of a signalling protein used in bacterial communication.

Bacteria can communicate with each other by extracellular signalling molecules called autoinducers. These molecules are often specific to a particular species of bacterium, and coordinate processes such as production of 'virulence factors' (which are involved in defence or attack). In the mid-1990s, AI-2, a general, nonspecific signalling molecule, was discovered. But AI-2's chemical identity proved elusive until its trace appeared in the electron-density map of its LuxP receptor, seen by Chen and colleagues using protein crystallography. This technique, then, can still be considered an analyti-

cal tool for identifying molecules that cannot be characterized by other methods. Moreover, Chen *et al.* find that an unusual element is involved — boron, to which a specific cellular role had not previously been assigned.

Boron is a trace mineral that is potentially toxic to all organisms. In the form of boric acid and borax, it is used as a pesticide and food preservative, and antiseptics containing boric acid are still a common source of accidents in young children. Boron is an essential mineral for green algae and higher plants, but no such role has yet been demonstrated for any other organism. The element might be involved in human brain function, alertness and cognitive performance, and in mineral metabolism. It might also reduce the symptoms or incidence of arthritis. But these are just possibilities: they don't help to establish that boron is an essential element in these circumstances, let alone define its cellular function. Chen and colleagues' discovery that boron is part of a signalling molecule in bacteria defines such a function.

The serendipitous characterization of the AI-2 autoinducer molecule is most intriguing. We live in an age of discovery fuelled by an explosion of gene-sequence information. The genetic make-up of organisms at all levels of the evolutionary ladder is being determined so rapidly that it is becoming increasingly difficult to keep up with the interpretations of the functions for gene products.

At present, these interpretations depend heavily on comparisons between genes that encode proteins of unknown function and those coding for proteins whose functions have been studied by classical biochemical and genetic methods. But it is unlikely that we will ever be 100% successful in identifying function by comparing amino-acid sequences. In part, this is because proteins from different organisms may be related by sequence but have very different functions. Sequence codes for the architecture of the protein, as well as the identity of an active or interaction site, so these two properties may not be preserved simultaneously. In addition, the identity of another molecule that interacts with a given protein cannot be deduced from its sequence.

The three-dimensional structure can be a

powerful ally when trying to identify the function of a protein whose sequence is unrelated to that of any other. At this level, however, surprises still occur. A sequence that is not related to that of a protein of known structure can nevertheless fold into an architecture that is already well characterized. In other words, the same three-dimensional protein fold can be derived from completely unrelated sequences. Moreover, although two proteins may be unrelated in sequence and structure, their active sites may be alike and identical in the job they carry out: a protein's function can be derived from more than one architecture.

Structure determination can also deliver a surprise when another molecule unexpectedly comes along for the ride. In these instances, the first assumption is that the bound molecule is specific for the protein and is therefore involved in what it does. That seems to be the case with the signalling

receptor LuxP, where the bound molecule is presumed to be the signalling molecule, AI-2. Chen *et al.* were able to establish the identity of AI-2 at the same time as seeing its interaction with the receptor.

The overall message then, is this. To provide a complete functional description of a gene product, information from many types of experiment has to be integrated. Hypothesis-driven research leads to clearly designed experiments that test well-defined ideas. Discovery-driven research is designed to find new correlations from a sea of information that is being generated for its own sake. The combination of the two should prove to be more powerful than either alone in identifying the cellular roles of undefined gene products. ■

Dagmar Ringe is at the Rosenstiel Basic Medical Sciences Research Center, Brandeis University, 415 South Street, Waltham, Massachusetts 02454-9110, USA.

e-mail: ringe@binah.cc.brandeis.edu

Biogeography

Big thinking

Brian A. Maurer

Since the demise of the dinosaurs, no land vertebrate has matched them for size. Why? The answer may lie in the particular conditions prevailing in the Cretaceous period.

How does evolution produce giant land vertebrates? One of the prerequisites for making such creatures appears to be the existence of big continents¹⁻³. Writing in *Proceedings of the National Academy of Sciences*, however, Burness *et al.*⁴ demonstrate that the largest animals also eat plants and are cold-blooded.

The authors have analysed data on the largest vertebrate herbivores and carnivores from 30 islands and continents. For each land mass, the authors identified either the largest living species of vertebrate land animal or the largest species that became extinct within the past 65,000 years. They show that, for a land mass of a given size, the largest warm-blooded herbivore was bigger than the largest warm-blooded carnivore. Data for cold-blooded land vertebrates are sparse, because so few of them are the largest vertebrate on a land mass. But, for a given land mass, the authors show that cold-blooded carnivores were larger than warm-blooded carnivores. Similarly, the three data points on cold-blooded herbivores suggest that they are larger than warm-blooded herbivores. Burness and colleagues then go on to evaluate the factors that may have been responsible for the evolution of large dinosaurs.

Like many ecological relationships, that between maximum body mass and land area is a power function⁴. That is, if M is the maximum body size of a group of terrestrial

vertebrates on a particular continent, and A is the area of that continent, then $M = kA^b$. Burness *et al.* found that the exponent b effectively equals 0.5, but that the constant k differs among warm- and cold-blooded herbivores and carnivores. The largest warm-blooded herbivore on any given land mass is roughly an order of magnitude larger than the biggest warm-blooded carnivore, if there is one. The few data that are available suggest a similar relationship for cold-blooded vertebrates.

Why should the largest carnivores be smaller than the largest herbivores? It is well established in ecology that the energy available to organisms declines by about 90% for each step up the food chain. So it is reasonable to expect that the largest size a carnivore can reach will be about 10% of the size of the largest herbivore on the same land mass.

Given a specific power function that describes the relationship between maximum body size and continental area for a given group of organisms, how does such a relationship arise? If maximum body mass is controlled in some manner by continent size, then continent size must be related to the minimum number of individuals in the geographical population of a large-bodied species. Each individual of a species of a given size requires a certain minimum amount of space within which to obtain resources and mates. This space, the organism's 'home range', is larger for animals of larger body

size⁵. The relationship between home range size, H , and body size is also a power function: $H = aM^z$. The total number of home ranges available for a species of a given size on a given continent will be the continent size divided by the size of that species' home range. After a little algebra, one can deduce that b (the exponent for the power function relating maximum body mass to continent size) should be the inverse of z (the exponent for the power function relating home range to body mass).

Testing whether $b = 1/z$ is true for the largest animals on islands and continents isn't entirely straightforward. Data to test this idea are currently available only for mammals, not reptiles or amphibians. There are no instances where the largest mammal on an isolated island or continent is less than about 100 g in mass². So although many mammals are smaller than 100 g, the test will be valid only for those greater than 100 g. Given these restrictions, Burness *et al.* analysed some previously published data⁵. When they fitted a power function to the sizes of large mammals as a function of their territory size, the resulting exponent was very close to 0.5, confirming the prediction. So large mammals are found only on large continents because small continents don't have enough space to maintain adequate populations of them.

Where do dinosaurs fit into the relationship between maximum body size and continent size? If dinosaurs were warm-blooded

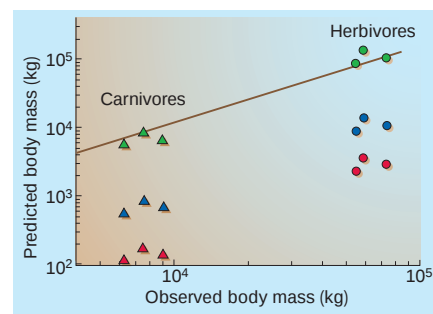


Figure 1 Why were the biggest dinosaurs so huge? Shown in red and blue are estimates of the sizes of several of the largest herbivorous (circles) and carnivorous (triangles) dinosaurs derived from modern relationships between continent size and maximum body size⁴. It is still not known whether dinosaurs were warm- or cold-blooded. Estimates made using relationships for warm-blooded modern vertebrates (red) are two orders of magnitude lower than the actual sizes (solid line); but those using relationships for cold-blooded vertebrates (blue) are still an order of magnitude lower. A possible explanation for the failure of these estimates, proposed by Burness *et al.*⁴, is that the productivity of terrestrial vegetation was ten times higher in the Cretaceous period, between 144 million and 65 million years ago, when dinosaurs reached their largest sizes. That vast extra productivity allowed animals to evolve that were an order of magnitude larger than those that modern continents can support (green symbols).

Daedalus

The analysis of fug

Last week Daedalus proposed a self-supporting Langmuir–Blodgett film. Its long-chain molecules, unlike those of normal films, are at right-angles to its own plane. DREADCO 'Langblofilm' may need its own porous support, but it lets small molecules percolate across it — in effect it is riddled with tiny holes. With an aluminium electrode layer evaporated on each side, and a voltage applied, liquid water molecules are driven across it by electro-osmosis. In this way it will remove water vapour from a room. As a wallpaper, it can pump water into the porous wall, from which it will evaporate and disperse.

Langblofilm is also ideal for removing liquid water from air, which holds it as humidity. Water users, from campers to horticulturalists and even to whole dry states, will welcome it not merely as a dehumidifier, but as a way of being independent of taps, rivers and reservoirs.

But Daedalus wants his film to do more than merely drying air. It would be ideal for removing smell. It could keep any kitchen sweet, replacing activated charcoal, air-fresheners, other domestic absorbents and air conditioners. In a living room it will pump away the human 'fug' of a dense gathering. Carbon dioxide is probably innocent. In spacecraft and nuclear submarines, human beings share a small volume of recycled air. Their sweat and trace emissions give a lot of trouble. The key molecules are probably small with high permittivity, ideal for Langblofilm. Its vast area and silent action should condense them on the wall and pump them outside without anybody even noticing.

DREADCO biochemists are now designing a small Langblofilm unit of higher voltage, to collect human 'fug' and concentrate it for research purposes. The result should be a liquid, mainly water but with the crucial smell molecules dissolved in it. Many social observers have claimed that a particular social scene has the 'wrong smell'. Emotions such as interest, fear and even panic are claimed to have smells of their own, detectable by dogs and other animals.

Non-human assemblies have their own characteristic 'fugs', as visitor to zoos have commented. So the study of fug might reveal a whole new biochemical way of looking at animal species, their emotions, and how these spread. Instinctive human sympathies and antipathies would be clarified, too. And now that military conflict occurs in caves and enclosed spaces, the study could have useful anti-social applications of its own. **David Jones**



Figure 2 Extra large: dinosaurs, such as these three sauropods, grew much bigger than later herbivores, like the mammoth, even on a land mass of comparable size.

they should more or less follow the relationship for warm-blooded organisms. But Burness *et al.* show that six species of dinosaurs for which estimates of body masses are available — three carnivores and three herbivores — were nearly two orders of magnitude larger than they should have been if they had been warm-blooded (Fig. 1). Using relationships for modern cold-blooded animals improved the prediction, but dinosaurs were still nearly an order of magnitude larger than predicted. The largest herbivorous and carnivorous dinosaurs lived in South America around 100 million years ago. The herbivore *Argentinosaurus huinculensis* weighed in at a massive 73,000 kg. The largest predator that may have fed upon this giant was *Giganotosaurus carolinii*, which tipped the scales at 9,000 kg.

So why were dinosaurs so much bigger than they would be if they had been alive more recently (Fig. 2)? The answer cannot be found by appealing to continent sizes. True, in the past, most of the continents were connected to form a megacontinent called Pangaea. But this land mass started breaking up nearly 180 million years ago, some 80 million years before the largest dinosaurs appeared⁶. By the time dinosaurs had achieved their largest sizes, the continents had separated and were roughly the same size as they are today.

Burness *et al.*⁴ suggest a possible explanation: during the Cretaceous period (about 144 million to 65 million years ago), concentrations of CO₂ in the atmosphere were up to ten times higher than they are today. An additional point, not made by the authors, is that the Cretaceous was also notable for the relatively small portion of the Earth's land mass that had cool temperate climates⁶. So during this time a larger proportion of the Earth's terrestrial environment was covered by warm temperate and tropical ecosystems than is the case today.

The higher atmospheric CO₂ and greater land mass in warm climates suggests that the amount of terrestrial vegetation available to support large animals was much greater during the Cretaceous, when the largest dinosaurs lived, than it is now. So it could have been this greater 'ecological size' of continents that allowed large dinosaurs to evolve. This idea might be tested by examining the sizes of modern large animals living in the tropical and temperate land masses of comparable size to see whether those in the tropics are bigger.

Dinosaurs present us with a puzzle. In many respects they seemed to be constructed like warm-blooded animals. Their posture indicated they were more active than living cold-blooded vertebrates. They apparently had extended parental care and complex mating rituals. Yet, ecologically, they filled continents as if they were cold-blooded. So what on the surface appears to be a case of convergent evolution between dinosaurs and modern vertebrates may in fact be the result of unique evolutionary events occurring in different ways at different times. Dinosaurs lived in a very different world from any modern animal, and may have interacted with their environment in ways that have no clear parallels among living land vertebrates. The more we study them, the more we get a glimpse into the complex workings of the evolutionary engine. ■

Brian A. Maurer is in the Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan 48823, USA.
e-mail: maurerb@msu.edu

1. Brown, J. H., Marquet, P. A. & Taper, M. L. *Am. Nat.* **130**, 1–17 (1993).
2. Marquet, P. A. & Taper, M. L. *Evol. Ecol.* **12**, 127–139 (1998).
3. Maurer, B. A. *et al.* *Evolution* **46**, 939–953 (1992).
4. Burness, G. P., Diamond, J. & Flannery, T. *Proc. Natl Acad. Sci. USA* **98**, 14518–14523 (2001).
5. Kelt, D. A. & Van Vuren, D. H. *Am. Nat.* **157**, 637–645 (2001).
6. Scotese, C. R. Paleomap Project, 1999; <http://www.scotese.com>

Obituary

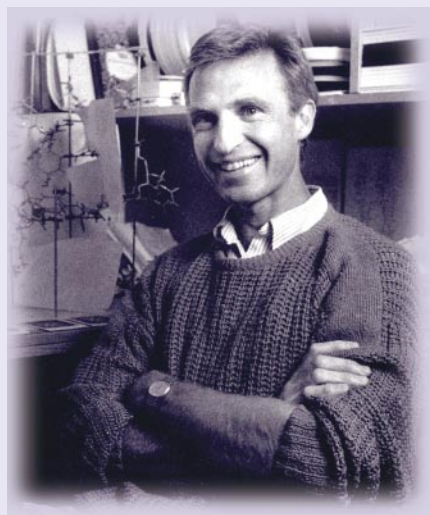
Don Craig Wiley (1944–2001)

The disappearance of Don Wiley, on 15 November 2001 in Memphis, Tennessee, completely mystified family and colleagues. The tragic denouement came with the discovery of his remains in a tributary of the Mississippi on 20 December. His death was ruled an accident.

Wiley was a crystallographer: this is the ultimate molecular biology. One of the twentieth century's greatest discoveries, the structure of DNA, has its roots in X-ray crystallography. In the same vein, the image of a class I MHC protein with its peptide cargo firmly in place will stand as a landmark Wiley discovery that forever changed the field of immunology. Discovered as a set of glycoproteins that are the target of transplant rejection, the products encoded by the MHC (major histocompatibility complex) are now known to alert the immune system to the presence of invaders, such as viruses, through the formation of a complex with snippets of the viral antigens, which is then displayed to patrolling T lymphocytes.

What makes Wiley's work so remarkable is the instant recognizability of the structures solved. Not only can immunologists draw on the back of an envelope the key features of an MHC product, but most virologists can sketch an outline of another of his triumphs — the influenza haemagglutinin protein, a component of the viral membrane or envelope. Asked by a colleague to expound on some biological question, Wiley demurred, saying: "I'm sorry, but I just don't understand anything in biology unless I know what it looks like." There is no better way to express crystallography's credo.

Wiley started his career as a crystallographer in the late 1960s at Harvard University, where he remained. Under the tutelage of William ('Colonel') Lipscomb, he began working on the enzyme aspartate transcarbamylase. Once independent, he turned his attention to the structure of membrane glycoproteins. Through a collaboration with John Skehel of the Medical Research Council labs at Mill Hill, London, a main focus became the structure of influenza haemagglutinins. This collaboration continued until Wiley's death, his stay at Mill Hill leading not only to the acquisition of an Aston Martin in British racing green (mothballed and never used much, as far as I recall), but also to one of the longest and most productive interactions of his career.



Crystallographer extraordinaire

The culmination was solving the structure of the full haemagglutinin trimer, which explained much about the structural basis of antigenic variation — the sequence of virus variants that evolve under selective pressure of the host's immune response against it. At the same time, it phrased the key questions for cell biologists studying the entry of the 'enveloped' viruses into animal cells. By 1980, the importance of a mildly acidic pH as the trigger required for virus entry had become part of cell biology lore. The structure of the influenza haemagglutinin trimer and its conformational alterations was the backdrop on which virus entry could now be projected. The drastic changes that accompany the transition of the neutral-pH to the low-pH form of haemagglutinin is a paradigm for the entry of many other enveloped viruses, and bears more generally on our understanding of membrane fusion.

Wiley's mapping work with defined elements of the haemagglutinin trimer provided an elegant demonstration of how one arm of the immune system — antibody defence — operates against viruses. Unveiling the other, the T-cell response, started with discovery of MHC 'restriction' in the mid-1970s. How recognition would be achieved in molecular terms led to a variety of remarkable and detailed proposals, all debunked in the end by the fruit of a long-standing collaboration between Wiley and Jack Strominger — the structure of the soluble portion of a class I MHC molecule known as HLA-A2. This

structure revealed the presence of a pocket, occupied by electron density not accounted for by the polypeptide backbone of the class I molecule itself. The pocket holds the antigenic peptide, which is recognized by the receptors of T cells only when firmly bound to the MHC product and presented at the surface of an infected cell. This structure was published in 1987; our concept of T-cell recognition has never been the same since.

Few protein structures explain a biological phenomenon in the way the class I MHC structure does. The biography of Dorothy Hodgkin recounts the first viewing of the insulin structure and the somewhat deflationary response of those invited to inspect it: formidable achievement though it was, insulin's structure did not reveal much about its mechanism of action. The class I MHC structure, by contrast, was a revelation. Wiley's accomplishments were recognized by a Lasker award and the Japan Prize (both shared with Strominger).

Wiley set high standards for himself and others, and shaped the field by training an impressive number of structural biologists, all of whom will do the Wiley legacy proud.

Don Wiley made crystallography fun to the non-crystallographer, too; anyone who heard him speak fell under the spell of his enthusiasm, and his ability to phrase complex biological questions in terms that opened them up for a crystallographic assault. Over the past few years, I would often go to his lab to see whether he was around for a chat — every instance was memorable. Don's enthusiasm for new discoveries would make even the most jaded sit up and take notice. Many will recall also the earlier days: the occasional retreat to the bar of the Hong Kong for scorpion bowls; and his irreverent treatment of the (in)famous Harvard fuddy-duddy establishment, laced with acerbic wit.

In the first half of December I passed through the corridor leading to his shuttered office to find the door open and a bewildered John Skehel at Don's desk, astounded at the level of organization left behind. Of course there is unfinished business, and we cannot say what scientific discoveries we have been deprived of by Don's death. But as a colleague, mentor and friend he is irreplaceable. He leaves behind his wife, Katrin Valgeirsdottir, four children and three grandchildren.

Hide L. Ploegh

Hide L. Ploegh is in the Department of Pathology, Harvard Medical School, 200 Longwood Avenue, Boston 02115, Massachusetts 02115, USA.
e-mail: ploegh@hms.harvard.edu

Photoreceptive net in the mammalian retina

This mesh of cells may explain how some blind mice can still tell day from night.

We have discovered an expansive photoreceptive 'net' in the mouse inner retina, visualized by using an antiserum against melanopsin, a likely photopigment^{1,2}. This immunoreactivity is evident in a subset of retinal ganglion cells that morphologically resemble those that project to the suprachiasmatic nucleus (SCN), the site of the primary circadian pacemaker^{3,4}. Our results indicate that this bilayered photoreceptive net is anatomically distinct from the rod and cone photoreceptors of the outer retina, and suggest that it may mediate non-visual photoreceptive tasks such as the regulation of circadian rhythms.

Most SCN-projecting retinal ganglion cells express melanopsin messenger RNA⁵. SCN-projecting retinal ganglion cells are also photosensitive⁶ and have spectral absorbance profiles that are characteristic of opsin-based photopigments⁶. The dimensions of the receptive fields of these cells are identical to those of their dendritic fields, indicating that photopigments that are capable of activating the depolarizing responses may reside in the dendrites (D. M. Berson *et al.*, personal communication).

We have found that melanopsin, the only opsin known to exist in retinal ganglion cells, is indeed expressed in the dendrites of this small subset of retinal ganglion cells (Fig. 1). Most striking are the plexuses formed by immunopositive dendrites in the outermost and innermost laminae of the inner plexiform layer (Fig. 2). These cells form an extensive network of opsin-containing dendrites, a photoreceptive apparatus previously overlooked in the mammalian retina.

The mammalian eye must complete two distinct photoreceptive tasks: vision and irradiance detection. Vision requires a fine

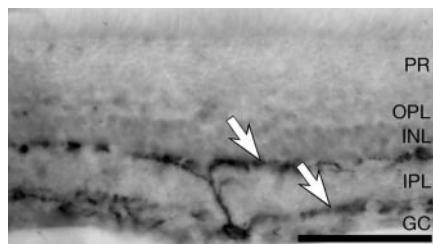


Figure 2 Cross-section of a mouse retina, with a single melanopsin-positive retinal ganglion cell. Note the two plexuses (arrows) of immunoreactive dendrites. GC, ganglion cells; IPL, inner plexiform layer; INL, inner nuclear layer; OPL, outer plexiform layer; PR, photoreceptors. Scale bar, 100 μ m.

array of 'narrow-capture' photoreceptive elements that allows fine spatial resolution; the photoreceptors (rods and cones) and downstream inner-retinal neurons involved in vision have been well characterized. By contrast, tasks such as photic regulation of circadian rhythms require the detection of changes in ambient irradiance, for which fine spatial resolution is not only unnecessary but may even confound the system. The retinal structures required for this 'broad-capture' photoreception have so far not been identified.

One strategy to determine irradiance levels from large sectors of visual space is to

sum or average the outputs of many narrow-capture photoreceptors. Alternatively, a distinct anatomical apparatus may fulfil the requirement for a broad-capture, integrating photoreceptive system. The expansive photoreceptive net described here within the inner retina could serve such a function. Moreover, this photoreceptive system could explain why photo-entrainment of circadian rhythms is abolished by bilateral eye removal, yet persists in mice that lack rods and cones⁷.

Ignacio Provencio, Mark D. Rollag, Ana Maria Castrucci

Department of Anatomy, Physiology and Genetics, and the Circadian Research Center, Uniformed Services University, Bethesda, Maryland 20814, USA
e-mail: iprovincio@usuhs.mil

1. Provencio, I. *et al.* *J. Neurosci.* **20**, 600–605 (2000).
2. Rollag, M. D., Provencio, I., Sugden, D. & Green, C. B. *Methods Enzymol.* **316**, 291–309 (2000).
3. Provencio, I., Cooper, H. M. & Foster, R. G. *J. Comp. Neurol.* **395**, 417–439 (1998).
4. Moore, R. Y., Speh, J. C. & Card, J. P. *J. Comp. Neurol.* **352**, 351–366 (1995).
5. Gooley, J. J., Lu, J., Chou, T. C., Scammell, T. E. & Saper, C. B. *Nature Neurosci.* **4**, 1165 (2001).
6. Berson, D. M., Dunn, F. A. & Takao, M. *Invest. Ophthalmol. Vis. Sci.* **42**, S113 Abstr. 613 (2001).
7. Freedman, M. S. *et al.* *Science* **284**, 502–504 (1999).

Competing financial interests: declared none.

Satellite imaging

Massive emissions of toxic gas in the Atlantic

Recurrent eruptions of toxic hydrogen sulphide gas in the waters along the Namibian coast off southwestern Africa have been considered to be local features with only limited ecosystem-scale consequences. But satellite remote sensing has revealed that these naturally occurring events are much more extensive and longer-lasting than previously suspected, and that the resultant hypoxia may last for much longer. The effects on the marine ecology and valuable coastal fisheries of this region are likely to be important.

In this richly productive coastal oceanic region^{1–3}, areas of seafloor hypoxia (< 0.5 ml O_2 l⁻¹), and even total anoxia, are extensive, and it is beneath these that toxic hydrogen sulphide can be generated within a metres-deep layer of diatom ooze. People living along the coast have become inured to the disagreeable smells and corrosive effects of sulphurous gases released from the sea, which kill a variety of nearshore fish

and invertebrates annually. This results in a fortuitous feast for seabirds, which converge on the milky turquoise-coloured patches of sulphide-infused waters to feed on floating casualties, and also for the local people, who eagerly harvest rock lobsters fleeing ashore from the noxious conditions in the water. However, the sheer magnitude of the extent, frequency and duration of these phenomena, as revealed by our recent satellite observations, was not realized.

For example, a quasi-true-colour image from the OrbView-2 SeaWiFS satellite for 18 March 2001 (Fig. 1a) shows an extended turquoise patch stretching more than 200 km along the Namib Desert coast. Ground personnel simultaneously recorded the hydrogen sulphide in water samples, together with observations of noxious sulphide odours, seabirds feeding on floating dead marine organisms, sharply anomalous dislocation of lobster distributions towards the shore, and so on, constituting evidence of intense hydrogen sulphide emissions in that zone. The concentration of dissolved oxygen was very low even in the surface 'milky' water (0.7 ml l⁻¹); the milky discoloration is created by the slurry of highly

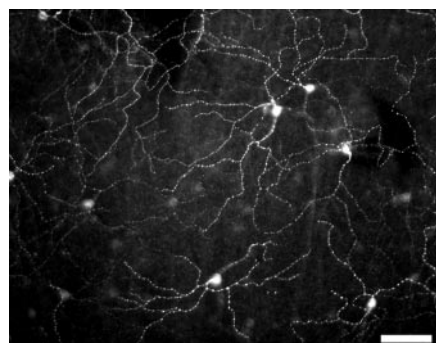


Figure 1 The photoreceptive net in the mouse inner retina. Immunofluorescent labelling of melanopsin-containing retinal ganglion cells on a flat mount of mouse retina reveals an extensive network of immunopositive dendrites. Scale bar, 100 μ m.

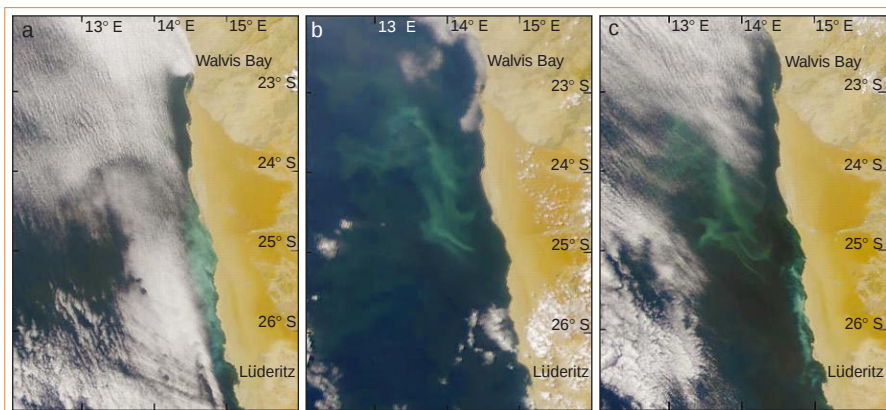


Figure 1 Sea-surface distribution of suspended sulphur granules. Quasi-true-colour images for the region at latitude 12° E– 16° E, longitude 22° S– 27° S, off south-central Namibia; images were generated from the OrbView-2 SeaWiFS satellite during March–April 2001. The milky turquoise coloration represents high concentrations of suspended sulphur granules. Data were collected on **a**, 18 March; **b**, 29 March; and **c**, 3 April 2001.

reflective precipitated microgranules of sulphur resulting from the oxidation of sulphide ions near the oxygenated sea surface.

In the days that followed, the feature was observed from satellite pictures to be advected northwards and offshore in the prevailing equatorward geostrophic current flow and offshore-directed surface-wind drift fields, which are characteristic of such eastern-ocean coastal upwelling systems⁴, to reach the position and configuration seen in Fig. 1b. In the image of 3 April (Fig. 1c), even while the earlier offshore feature continues to maintain a coherent identity, another totally new hydrogen sulphide emission event is seen to have started abruptly within the coastal upwelling zone north of Lüderitz.

Once the earlier feature had separated from the coast, it took on an appearance superficially similar to that in satellite views of very large coccolithophore blooms⁵. However, the continuity of the feature in the available satellite views and the advective context, as inferred from the evolution of satellite-sensed ocean surface temperature, make it clear that the offshore feature is the same as that previously observed as an intense episode of sulphide emission nearer the coast. The zone of intense coloration in later images (Fig. 1b) extends over an expanse of total sea surface greater than 20,000 km².

Such episodes have been recorded somewhere along the coast of Namibia more often than not during the recent months since the satellite observational capability was recognized. Where verification on the ground has been available, the early stages of these events have likewise been confirmed to be associated with sulphide emissions. In addition to its directly toxic effects, hydrogen sulphide strips dissolved oxygen from the water column, leaving behind a subsurface hypoxia that is manifested visibly as the sulphur-infused sea surface. It is this subsurface hypoxia,

which might endure well beyond the period when the toxic sulphide gas is present, that could pose the principal ecological problem. An event, first clearly noted on 17 May 2001, was still plainly visible as late as 6 June 2001.

Scarla J. Weeks*, Bronwen Currie†, Andrew Bakun‡

**Ocean Space Ltd, and †IRD, IDYLE Project, Oceanography Department, University of Cape Town, Rondebosch 7701, South Africa*
e-mail: oceanspace@icon.co.za

‡National Marine Research and Information Center, Swakopmund, Namibia

1. Bailey, G. W., Beyers, C. J. deB. & Lipschitz, S. R. *S. Afr. J. Mar. Sci.* **3**, 197–214 (1985).
2. Mas-Riera, J., Lombarte, A., Gordo, A. & Macpherson, E. *Marine Biol.* **104**, 175–182 (1990).
3. Hamukuaya, H., O'Toole, M. J. & Woodhead, P. M. J. *S. Afr. J. Mar. Sci.* **19**, 57–59 (1998).
4. Wooster, W. S. & Reid, J. L. in *The Sea Vol. 2* (ed. Hill, M. N.) 253–280 (Interscience, New York, 1963).
5. Tyrrell, T., Holligan, P. M. & Mobley, C. D. *J. Geophys. Res.* **104**, 3223–3241 (1999).

Competing financial interests: declared none.

Biomechanics

Dinosaur locomotion from a new trackway

Ardley Quarry in Oxfordshire, UK, contains one of the most extensive dinosaur-trackway sites in the world, with individual trackways extending for up to 180 metres. We have discovered a unique dual-gauge trackway from a bipedal theropod dinosaur from the Middle Jurassic in this locality, which indicates that these large theropods were able to run and that they used different hindlimb postures for walking and running. Our findings have implications for the biomechanics and evolution of theropod locomotion.

The Ardley trackways are preserved on a single horizon of the Middle Bathonian

(163 million years old)¹ white-limestone formation and include those from large theropods and at least two types of sauropod dinosaur. Three of the trackways (numbered 13, 29 and 80; J.J.D., unpublished observations) comprise large tridactyl ('three-toed') prints, with narrow claw impressions typical of theropod dinosaurs (Fig. 1). All three trackways (except for a section of track 13) are wide-gauge, with the prints indicating that the hind feet were placed sequentially in a zig-zag arrangement separated widely from the midline (Fig. 1a). Stride length (2.70 m) and pace angulation (117° – 132° ; Fig. 1) remain constant throughout the wide-gauge tracks. This pattern contrasts strongly with the more usual narrow-gauge form of theropod trackway^{2,3} in which pace-angulation values range between 160° and 170° : the hind-foot impressions are located close to the midline of the trackway.

Trackway 13 is unique because one section shows the theropod gait switching from wide to narrow gauge (the stride length increases to 5.65 m and pace angulation increases to 173° ; Fig. 2). The orientation of the feet also changes from positive rotation (toes directed forwards and inwards; Fig. 1a) during wide-gauge locomotion, to a small negative rotation (toes directed slightly outwards) during the narrow-gauge phase (Fig. 1b).

Estimating the speed of movement of the track-maker requires information on

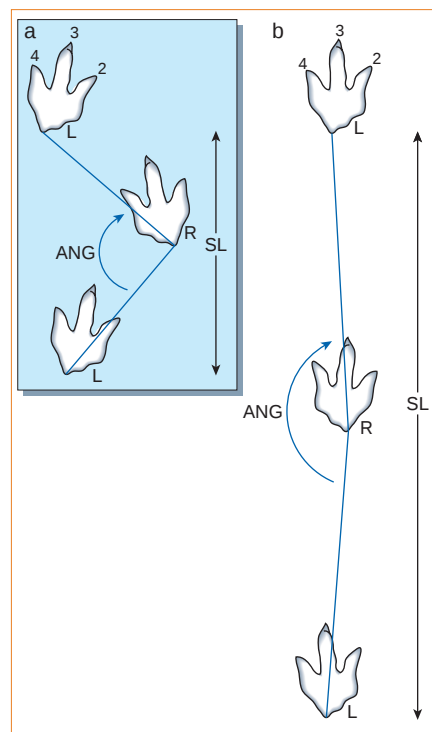


Figure 1 The dual-gait features of theropod trackway 13 from Ardley Quarry, Oxfordshire, UK. **a**, Wide-gauge section; **b**, narrow-gauge section. Digits are numbered and footprints are identified as left (L) and right (R); ANG, pace angulation; SL, stride length.

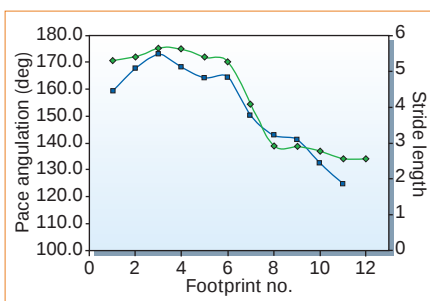


Figure 2 Deceleration phase of the 'running' section of trackway 13. Blue, pace angulation; green, stride length. Prints preceding this section are missing but global-positioning satellite data (J.J.D., unpublished observations) indicate that this section is continuous with the rest of trackway 13.

hip height (h) and stride length^{3,4}. Hip height can be calculated from foot length³ (which equals the length of the middle toe, 'digit III'). For trackway 13, this gives an h value of 1.93 m, giving estimated speeds of 6.8 km h⁻¹ and 29.2 km h⁻¹ for the wide- and narrow-gauge segments, respectively.

Placing the hind feet close to, or on, the midline⁵, optimizes locomotor efficiency by maximizing the effective stride length and reducing the energy lost through lateral displacement of the centre of gravity. The Ardley trackways therefore have important implications for the locomotor styles of large theropod dinosaurs. The simultaneous alteration of stride length and pace angulation (Fig. 2), foot orientation and estimated speeds all suggest that the large Ardley theropod used the wide-gauge gait while walking but was capable of a 'gear change' to maximize running ability by employing the more efficient, narrow-gauge gait.

Greater stability was probably achieved during walking by shortening the stride length, placing the hind feet further from the midline and rotating the feet inwards. Although providing more stability, the wider spacing of the hindlimbs would decrease the maximum stride length and therefore the attainable speed, so faster locomotion would have called for a change to the narrow-gauge gait.

Trackway evidence of the running ability of large bipedal dinosaurs is scant⁶. Estimates of maximum speeds obtained from biomechanical studies of skeletal material are controversial, particularly for large theropods³. The increase in stride length evident in trackway 13 indicates that at least some large theropods were capable of running at speeds close to previous maximum estimates.

Other issues concerning locomotion in large theropods remain to be resolved. The narrow-gauge (running) section of trackway 13 is only 35 m long and leaves unanswered the question of how long a large theropod could sustain a running gait. It is also uncertain whether the wide-gauge gait

was habitually adopted by large theropods while walking, or only when the ground was very soft or unstable. Furthermore, the anatomical correlations between leg and hip anatomy associated with the adoption of wide- and narrow-gauge gaits are not yet known and we are therefore unable to determine the phylogenetic distribution of these locomotor styles. Nevertheless, the Ardley trackways offer new insight into dinosaur locomotor capacity and will stimulate enquiry into the evolution and biomechanics of large theropod dinosaurs.

Julia J. Day*, David B. Norman*, Paul Upchurch*, H. Philip Powell†

*Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK

e-mail: jday00@esc.cam.ac.uk

†Oxford University Museum of Natural History, Parks Road, Oxford OX1 3PW, UK

1. Cope, J. C. W. *et al.* *Geol. Soc. Spec. Rep.* **15** (1980).
2. Padian, K. & Olsen, P. E. *Copeia* **1984**, 662–671 (1984).
3. Thulborn, R. A. *Dinosaur Tracks* (Chapman & Hall, London, 1990).
4. Alexander, R. M. *Nature* **261**, 129–130 (1986).
5. Alexander, R. M. *Zool. J. Linn. Soc.* **83**, 1–25 (1986).
6. Thulborn, R. A. *Nature* **292**, 273–274 (1981).

Fish physiology

Dogfish hair cells sense hydrostatic pressure

Many marine invertebrates and fish respond to hydrostatic pressure in order to regulate their depth and synchronize their behaviour to tidal cycles^{1–4}. Here we investigate the effect of hydrostatic pressure on the vestibular hair cells located in the labyrinth of the dogfish *Scyliorhinus canicula*, and find that it modulates their spontaneous activity and response to angular acceleration. This may explain not only the low resting activity of vertebrate hair cells but also how fish that do not have swim bladders can sense hydrostatic cues.

How a hydrostatic-pressure-sensing mechanism could exist in animals that have no gas compartment remained a mystery⁵ until the discovery of a piston mechanism in crabs⁶, in which differential compression of the outer cuticle and the internal fluid of thread hairs in the balancing system leads to nanometre-level displacements that are sensed by mechanoreceptors⁶. Cells in culture may respond to tiny changes in hydrostatic pressure¹ by altering their division or in other ways, although isolated cochlear hair cells show no significant length or transmembrane-voltage changes in response to increased hydrostatic pressure⁷.

We made extracellular recordings from

vestibular type-II hair cells in horizontal canal units in isolated elasmobranch labyrinths, a well-known model system⁸, from seven dogfish to test the sensitivity of these cells to hydrostatic pressure. Nerve spikes were conventionally amplified, filtered and recorded. A computer-controlled pressure regulator produced a test sequence of 1 hour at atmospheric pressure (taken as zero bar) followed by 1 hour at 14 kilopascals (kPa; 0.14 bar) and four sinusoidal pressure cycles of a further 30-kPa amplitude and 15-min period. To stimulate the vestibular system, the preparation was mounted on an Inland DC Servomotor system inside a pressure chamber. A bout of 64 cycles of oscillation around the vertical axis at 1 Hz and 70 deg amplitude was supplied every 2 min. Single units (sets of nerve-spike activity from single neurons) or small groups of units were averaged online or counted offline. Temperature changes were restricted to less than 0.5 °C per hour.

Units responded unidirectionally to oscillation at 1 Hz. Resting and oscillation-derived spike densities (Fig. 1) increased significantly (Mann–Whitney *U*-test) after the pressure step. During, but not preceding, the imposed hydrostatic-pressure cycles, related cyclical changes in spike frequency were confirmed by using autocorrelation and maximum-entropy spectral analysis⁹. There was a positive relationship between spike frequency and hydrostatic pressure but not between spike frequency and rate of pressure change.

The changes in resting and angular-acceleration-derived activity after a change in hydrostatic pressure could be part of a depth-sensing mechanism in fish and may explain the signalling value of low resting activity in hair cells. It is likely that differential compression leads to nanometre-scale strains that activate the hair cells directly.

This finding has implications for transduction mechanisms in a variety of hair-cell types, including cochlear hair cells, in

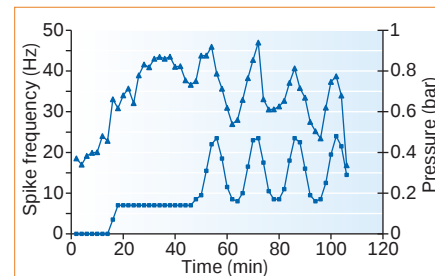


Figure 1 Average extracellular spike frequencies from a small set of units from the dogfish right vestibular system during bouts of 64 oscillation cycles (top trace) and hydrostatic pressure (bottom trace) recorded every 2 min. Oscillation-derived activity increases significantly after the change in pressure and shows cyclical changes with the same 15-min period as the imposed pressure cycles.

which there are complex electromechanical mechanisms, and fish lateral-line hair cells. Although no analogous structure to the long, narrow thread hair in the crab is known in vertebrate hair cells, which do not have piston mechanisms, the greater numbers of hair cells in these organisms may help to increase signal-to-noise ratios.

Peter J. Fraser, Richard L. Shelmerdine

Department of Zoology, University of Aberdeen,
Tillydrone Avenue, Aberdeen AB24 2TZ, UK
e-mail: p.fraser@abdn.ac.uk

1. Macdonald, A. C. & Fraser, P. J. *Comp. Biochem. Physiol. A* **122**, 13–36 (1999).
2. Hardy, A. C. & Bainbridge, R. *Nature* **168**, 327–328 (1951).
3. Blaxter, J. H. S. in *Environmental Physiology of Fishes* (ed. Ali, M. A.) 369–386 (Plenum, New York, 1980).
4. Pavlov, D. S., Sadvovskii, R. V., Kostin, V. V. & Lupandin, A. I. *J. Fish Biol.* **57**, 69–81 (2000).
5. Digby, P. S. & Leach, B. *Nature* **191**, 366–368 (1961).
6. Fraser, P. J. & Macdonald, A. G. *Nature* **371**, 383–384 (1994).
7. Conradi, P. & Ulfendahl, M. *Oto-Rhino-Laryngology* **61**, 57–62 (1999).
8. Lowenstein, O. & Compton, G. J. *Proc. R. Soc. Lond. B* **202**, 313–338 (1978).
9. Dowse, H. B. & Ringo, J. M. *Biol. Rhythm Res.* **25**, 1–13 (1994).

Plankton blooms

Lysogeny in marine *Synechococcus*

Viral infection of bacteria can be lytic, causing destruction of the host cell, or lysogenic, in which the viral genome is instead stably maintained as a prophage within its host¹. Here we show that lysogeny occurs in natural populations of an autotrophic picoplankton (*Synechococcus*) and that there is a seasonal pattern to this interaction. Because lysogeny confers immunity to infection by related viruses¹, this process may account for the resistance to viral infection seen in common forms of autotrophic picoplankton².

We undertook a seasonal study in Tampa Bay, Florida, of prophage induction in cyanobacteria over the year ending in October 2000 to find out whether lysogeny occurs in natural *Synechococcus* populations and, if so, how it is affected by changing environmental conditions.

Cyanophage abundance was determined by serial dilution on microtitre plates with addition of the host organism *Synechococcus* st. CCMP 1334 (WH7803, DC2); data were processed by using a most-probable-number (MPN) program³. SYBR gold staining followed by epifluorescence microscopy⁴ confirmed the presence of viral particles in wells containing lysed cells. To determine whether lysogens were present, we assayed water samples for prophage induction using mitomycin C as the inducing agent. A viral-reduction technique⁵ was used in the preparation of samples to

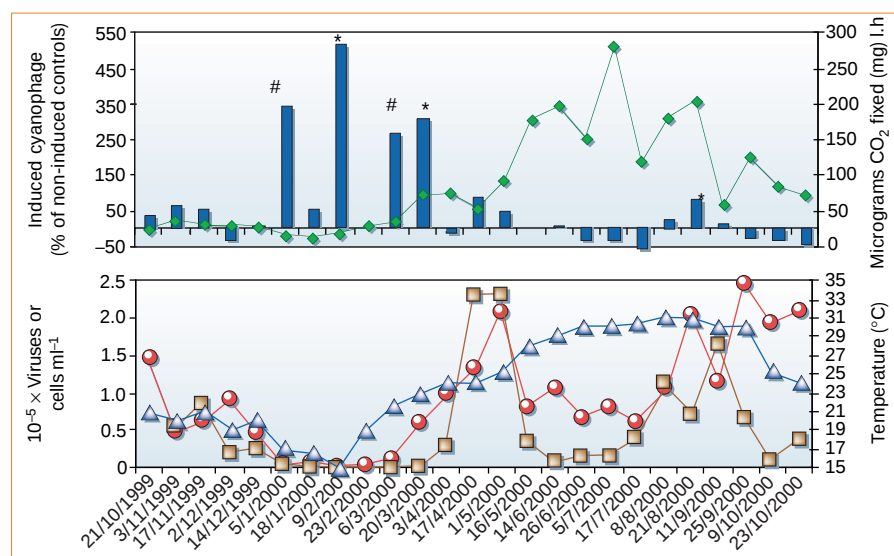


Figure 1 Seasonal induction of prophage in natural *Synechococcus* populations. Top, variation in cyanophage induction (blue bars) compared with primary productivity of *Synechococcus* (green line) for the year ending in October 2000. l.h., litre hour. Bottom, corresponding values for *Synechococcus* abundance (red), cyanophage counts (brown) and temperature (blue) over the same period. Prophage-induction results are expressed as a percentage change in treatment compared with control (asterisks and hash symbols indicate statistical significance at the 95% and 90% confidence interval, respectively; the significance of each induction event was determined by comparison of treatment and control levels of cyanophage by paired *t*-test in three pseudoreplicates of each sample).

increase the sensitivity of the assay.

A statistically significant amount of prophage induction in natural populations of *Synechococcus* was revealed on six occasions in response to exposure to mitomycin C (Fig. 1). Prophage induction, measured as the percentage change over the control cyanophage level, was inversely correlated with cyanobacterial abundance and primary productivity ($r = -0.502$, $P = 0.0123$ and $r = -0.4109$, $P = 0.0461$, respectively; correlations were determined by multiple-regression analysis of arcsine-transformed percentage data and all other measured parameters).

Induction occurred primarily during the late winter months, during times of reduced host abundance (Fig. 1). One induction event was also observed in late August, which preceded a secondary autumn bloom in *Synechococcus*, indicating that prophage induction in *Synechococcus* was not simply an artefact resulting from lower cyanophage abundance during winter months.

A seasonal pattern is consistent with the occurrence of lysogeny at times of low host availability, resource limitation or adverse environmental conditions in order to ensure viral survival¹. The limited lysogeny that occurred during the summer months may also be explained by greater exposure to ultraviolet light and/or higher temperatures having already caused induction of many of the prophage.

The MPN method of cyanophage detection has a precision that is comparable to that of other methods of viral enumeration⁶. However, it will only detect a subset of the lysogenic *Synechococcus* population (that is, phage infective for *Synechococcus* st.

CCMP 1334) and the total number of inducible cyanophages must be higher. Our values are therefore a conservative estimate of the actual number of lysogenic *Synechococcus* present.

Viruses are abundant in the marine environment⁷, infecting both heterotrophic and autotrophic microbial populations⁸, modulating microbial production and, in some cases, terminating algal blooms⁹. Cyanobacteria can be resistant to lytic infection by co-occurring cyanophage⁸, and *Synechococcus* strains may also be less sensitive than larger species of phytoplankton to the photosynthetic inhibition caused by the viral fraction of sea water¹⁰. We propose that these observations may be explained if homo-immunity is conferred by lysogeny.

**L. McDaniel, L. A. Houchin,
S. J. Williamson, J. H. Paul**

University of South Florida, College of Marine
Science, St Petersburg, Florida 33701, USA
e-mail: jpaul@seas.marine.usf.edu

1. Ackermann, H. W. & DuBow, M. S. *Viruses of Prokaryotes* (CRC, Boca Raton, Florida, 1987).
2. Waterbury, J. B. & Valois, F. W. *Appl. Environ. Microbiol.* **59**, 3393–3399 (1993).
3. Suttle, C. A. & Chan, A. M. *Mar. Ecol. Prog. Ser.* **92**, 99–109 (1993).
4. Noble, R. T. & Fuhrman, J. A. *Aquat. Microb. Ecol.* **14**, 113–118 (1998).
5. Weinbauer, M. G. & Suttle, C. A. *Appl. Environ. Microbiol.* **62**, 4374–4380 (1996).
6. Cottrell, M. T. & Suttle, C. A. *Limnol. Oceanogr.* **40**, 730–739 (1995).
7. Bergh, Ø., Børsheim, K. Y., Bratbak, G. & Heldal, M. *Nature* **340**, 467–468 (1989).
8. Wommack, K. E. & Colwell, R. R. *Microbiol. Mol. Biol. Rev.* **64**, 69–114 (2000).
9. Bratbak, G., Egge, J. K. & Heldal, M. *Mar. Ecol. Prog. Ser.* **93**, 39–48 (1993).
10. Suttle, C. A. *Mar. Ecol. Prog. Ser.* **87**, 105–112 (1992).

Competing financial interests: declared none.

Genome sequence of the plant pathogen *Ralstonia solanacearum*

M. Salanoubat^{*†}, S. Genin[‡], F. Artiguenave^{*§}, J. Gouzy[‡], S. Mangenot^{*}, M. Arlat[‡], A. Billault^{||}, P. Brottier^{*}, J. C. Camus[‡], L. Cattolico^{*}, M. Chandler[¶], N. Choise[#], C. Claudel-Renard^{*}, S. Cunac[‡], N. Demange^{*}, C. Gaspin^{**}, M. Lavie[‡], A. Moisan^{**}, C. Robert^{*}, W. Saurin^{*§}, T. Schiex^{**}, P. Siguier[¶], P. Thébaud[‡], M. Whalen^{‡§}, P. Wincker^{*}, M. Levy^{*}, J. Weissenbach^{*} & C. A. Boucher[‡]

^{*} Genoscope and CNRS UMR-8030, 2 rue Gaston Crémieux, CP5706, 91057 Evry Cedex, France

[‡] Laboratoire de Biologie Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France

^{||} Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France

[¶] LMGM CNRS, 118 Route de Narbonne, F 31062 Toulouse Cedex, France

[#] Genoscope and INRA URGV, 2 rue Gaston Crémieux, CP5706, 91057 Evry Cedex, France

^{*} Laboratoire de Genétique Cellulaire, INRA, BP27, 31326 Castanet-Tolosan Cedex, France

^{**} Unité de Biométrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex, France

[†] These authors contributed equally to this work

***Ralstonia solanacearum* is a devastating, soil-borne plant pathogen with a global distribution and an unusually wide host range. It is a model system for the dissection of molecular determinants governing pathogenicity. We present here the complete genome sequence and its analysis of strain GMI1000. The 5.8-megabase (Mb) genome is organized into two replicons: a 3.7-Mb chromosome and a 2.1-Mb megaplasmid. Both replicons have a mosaic structure providing evidence for the acquisition of genes through horizontal gene transfer. Regions containing genetically mobile elements associated with the percentage of G+C bias may have an important function in genome evolution. The genome encodes many proteins potentially associated with a role in pathogenicity. In particular, many putative attachment factors were identified. The complete repertoire of type III secreted effector proteins can be studied. Over 40 candidates were identified. Comparison with other genomes suggests that bacterial plant pathogens and animal pathogens harbour distinct arrays of specialized type III-dependent effectors.**

Studies addressing the molecular determinants of bacterial pathogenicity towards plants have concentrated on a limited number of bacterial species, representing the taxonomic diversity of principal Gram-negative plant pathogens that produce the most common diseases. Among these, *R. solanacearum*^{1,2} has unique and significant features. It is a soil-borne pathogen that naturally infects roots. It exhibits a strong and tissue-specific tropism within the host, specifically invading, and highly multiplying in, the xylem vessels. In addition, with over 200 host species belonging to more than 50 botanical families³, this bacterium has an unusually wide host range, and offers a unique opportunity for the analysis of a strong and generalized array of virulence factors. *Ralstonia solanacearum* has been studied intensively both biochemically and genetically, and has long been recognized as a model system for the analysis of pathogenicity⁴. It is well adapted to life in soil in the absence of host plants⁵, thereby providing a good system to investigate functions governing adaptation to such an ecological niche. *Ralstonia solanacearum* is a β -proteobacterium and thus belongs to a group of bacteria whose genomic organization is still poorly characterized. The only other representative of this group for which the complete genome sequence has been determined and annotated is *Neisseria meningitidis*^{6,7}.

To date, the genome of only one plant pathogenic bacterium, *Xylella fastidiosa*, has been sequenced completely⁸. Here we report the complete nucleotide sequence of the genome of *R. solanacearum* strain GMI1000, a race 1 strain isolated from tomato. Notably, GMI1000 is pathogenic on the model plant *Arabidopsis thaliana*, the genome of which has been entirely sequenced⁹, thereby facilitating studies of host response. We provide an integrative analysis of the predicted functions encoded in this organism with special emphasis on pathogenicity. Our analysis of the genome sequence provides

clues to the evolution of pathogenicity functions. The genome sequence of *R. solanacearum* is a first step towards an exhaustive functional analysis of pathogenicity determinants in this pathogen.

A bipartite genome structure

After sequencing by the whole-genome random sequencing method, we assembled the *R. solanacearum* genome into two circular molecules: a large replicon of 3,716,413 bp and a smaller 2,094,509-bp replicon, yielding a total genome size of 5,810,922 bp. The two molecules have an almost identical G+C content (67.04% and 66.86% for the large and small replicon, respectively). These two molecules correspond to the two bands that are visualized on pulsed-field gel electrophoresis (see Supplementary Information Fig. 1). The genome encodes a total of 5,129 predicted proteins, and the general features are shown in Table 1.

Analysis of the sequence establishes that the smaller replicon corresponds to the previously described megaplasmid, as it encodes *hrp* genes that were previously localized on this replicon¹⁰. Such a bipartite genome structure seems to be a characteristic of *R. solanacearum*, as a megaplasmid has been detected in most of the

Table 1 General features of the *R. solanacearum* strain GMI1000 genome

Genome feature	Chromosome	Megaplasmid	Genome
Length (bp)	3,716,413	2,094,509	5,810,922
G+C ratio	67.04%	66.86%	66.97%
Protein-coding regions	87.8%	86.5%	87.3%
tRNAs	55	3	58
Ribosomal RNA operons	3	1	4
Protein-coding genes	3,448	1,681	5,129
Average length of protein-coding genes (bp)	946	1,077	989
Genes with functional assignment	1,609	652	2,261
Orphan genes*	12.6%	18.7%	14.6%
Regulatory genes*	7.2%	9.6%	8.0%
Insertion sequences and phage sequences†	3.4%	2.5%	3.1%

* Percentage of total protein-coding genes.

† Percentage of replicon size.

§ Present addresses: Genomining, 93–95 rue Henri Rochefort, 91000 Evry, France (F.A., W.S.); Department of Biology, San Francisco State University, 1600 Holloway Avenue, San Francisco, California 94132, USA (M.W.).

strains belonging to this species¹¹. As no derivative of strain GMI1000 in which the megaplasmid is deleted has been obtained, the status of this replicon as a plasmid is still an open question. The distribution of genes between the two replicons provides some insights into this question.

The origin of replication of the large replicon, as identified by G+C skew analysis¹², has features that are characteristic of a chromosomal origin of replication in bacteria (see Supplementary Information Fig. 2). It is flanked by the *rnpA* gene on one side, and the *dnaA*, *dnaN*, *gyrAB* genes on the other. It also harbours a single consensus DnaA-binding box (TTATCCACAAA), the first nucleotide of which was arbitrarily chosen as the origin for the nucleotide numbering of the large replicon. Furthermore, the large replicon encodes a complete set of essential housekeeping genes including genes required for: (1) DNA replication, DNA repair and cell division; (2) transcription; and (3) translation. This last set includes all the ribosomal proteins, 3 complete ribosomal DNA loci, and 55 identified transfer RNAs, allowing recognition of all possible codons. Finally, all the essential genes required for purine and pyrimidine biosynthesis are located on the large replicon. Therefore, this replicon encodes all of the basic mechanisms required for the survival of the bacterium, and is referred to here as the 'chromosome'. Accordingly, the smaller replicon may be a dispensable genetic element. The putative origin of replication of this replicon, as predicted by G+C skew analysis, has characteristics of plasmid-borne *ori* loci. It is flanked by the *repA* gene and by at least 14 repetitions of a conserved motif (consensus G/CCGTACCCG/ATTCTGCG) that may be RepA-binding boxes. Therefore, the smaller replicon appears to be a megaplasmid. The first nucleotide of the most upstream RepA-binding box located in the intergenic region preceding *repA* has been arbitrarily chosen as the origin for nucleotide numbering of the megaplasmid.

The megaplasmid carries several metabolically essential genes that are also present on the chromosome. These include a complete copy of a rDNA locus with 2 tRNA genes, a gene coding for the α -subunit of DNA polymerase III, and a gene for the protein elongation factor G. Moreover, several enzymes controlling primary metabolism, including amino acid and cofactor biosynthesis, are encoded on the megaplasmid with no counterpart on the chromosome. As a consequence, we predict that a megaplasmid-deleted derivative of strain GMI1000 will be auxotrophic for several metabolites, a status similar to that reported for certain internal deletion mutants of the megaplasmid¹³.

Analysis of the genes present on the megaplasmid suggests that this replicon has a significant function in overall fitness and adaptation of the bacterium to various environmental conditions. As mentioned previously, the megaplasmid carries all of the *hrp* genes that are required to cause disease on plants, a trait that allows the bacterium to colonize a rather exclusive ecological niche. The megaplasmid also encodes the constituents of the flagellum and most of the genes governing exopolysaccharide synthesis. In addition, this replicon carries 315 out of the 748 genes of unknown function, a proportion significantly biased in favour of the megaplasmic ($P = 4 \times 10^{-7}$). On the other hand, a significant bias in favour of the chromosome is observed for the *R. solanacearum* genes that are shared with other bacteria (see Supplementary Information Fig. 3).

Mosaic structure of the genome

The gene prediction software FrameD (<http://www.toulouse.inra.fr/FrameD.html>) using the probabilistic model constructed on previously characterized *R. solanacearum* genes, led to a clear-cut prediction of genes in more than 90% of the genome. Using this matrix, a significant portion of the genome (7%) was predicted as non-coding (in regions spanning over 1 kb) although in some instances, BLASTX analysis revealed significant similarities in these regions with known proteins. On the basis of these homo-

logies, an alternative matrix was constructed and used for gene prediction in such regions that we designated alternative codon-usage regions (ACURs). When analysed for base composition, most ACURs, but not all, differ significantly from the average 67% G+C content found for the entire genome, with variations ranging from 50% to 70% G+C content. In addition, codon usage in these regions differs significantly from codon usage in the rest of the genome (Supplementary Information Table 1 and Fig. 4). Furthermore, ACURs were often associated with mobile genetic elements. In 44 out of 93 ACURs, a prophage, insertion sequence or part of an insertion sequence occurred either encoded directly within the ACUR or within the 1-kb flanking region. The strongly biased distribution of genetically mobile elements with ACURs ($P = 1.7 \times 10^{-5}$) suggests that ACURs may have been acquired through horizontal gene transfer, consistent with the propensity of *R. solanacearum* to take up and recombine exogenous DNA through natural transformation¹⁴.

In addition to ACURs, strain GMI1000 harbours several other elements that may have a function in genetic instability and rapid evolution of the genome (Fig. 1). Throughout the genome, there are at least 118 copies of complete or truncated insertion sequences representing 17 distinct elements belonging to 7 families. Three of these insertion sequences have been identified previously in *R. solanacearum* (ISRso1, IS1421 and IS1021) (<http://www-IS.biotoul.fr>) and we have called the remaining 14 ISRso5–ISRso18. There is a preponderance of IS3 (22 copies) and IS5 (27 copies) family members. In addition, there are two, presumably non-autonomous, derivatives of ISRso1, composed uniquely of the terminal sequences separated by an identical 117-bp DNA segment. This is similar to the RUP elements observed in *Streptococcus pneumoniae*¹⁵. At least 4 possibly defective prophages were found on the chromosome together with a conjugative transposon located between positions 2,781,738–2,825,808 (Fig. 1). This transposon is related to the 55-kb transposon Tn4371 from *Ralstonia metallidurans*¹⁶. The *R. solanacearum* conjugative transposon includes a set of *tra* and *trb* genes for conjugation as well as an integrase (RS00926); however, the genes coding for biphenyl resistance in Tn4371 are replaced by a group of genes with undefined functions. There are several loci encoding Rhs and Vgr-related elements^{17,18}, found to be recombinational hotspots in *Escherichia coli*. Of note, ACURs and mobile genetic elements are not distributed evenly on the genome but are often clustered on both replicons (Figs 1 and 2).

The *R. solanacearum* sequence has a mosaic structure containing numerous elements signalling the potential for evolution. Genomic rearrangements have already been reported to occur naturally in this bacterium^{13,19} and are further exemplified by the almost perfect tandem duplication of a 31-kb stretch of DNA on the megaplasmid (positions 1,648,630–1,710,832). The present genome sequence may represent a single snapshot of a structure that is variable from isolate to isolate and within derivatives from the same isolate.

Candidate genes responsible for pathogenesis

Apart from the virulence genes already described in *R. solanacearum*, a series of new genes putatively involved in pathogenicity were identified (Table 2; see Supplementary Information Table 2 for a complete list). These include genes coding for additional hydrolytic enzymes involved in the degradation of plant cell walls, and genes required for the production of the plant hormones auxin and ethylene or for the degradation of the plant signalling molecules ethylene and salicylic acid—a mediator of the plant systemic acquired resistance. Ten genes were predicted to code for proteins involved in resistance to oxidative stress. These 10 genes may be involved in the detoxification of the active oxygen species produced by infected plants, molecules reportedly representing a first line of defence against pathogen invasion²⁰. Genes involved in the production of toxins or antibiotics were also identified: these include six

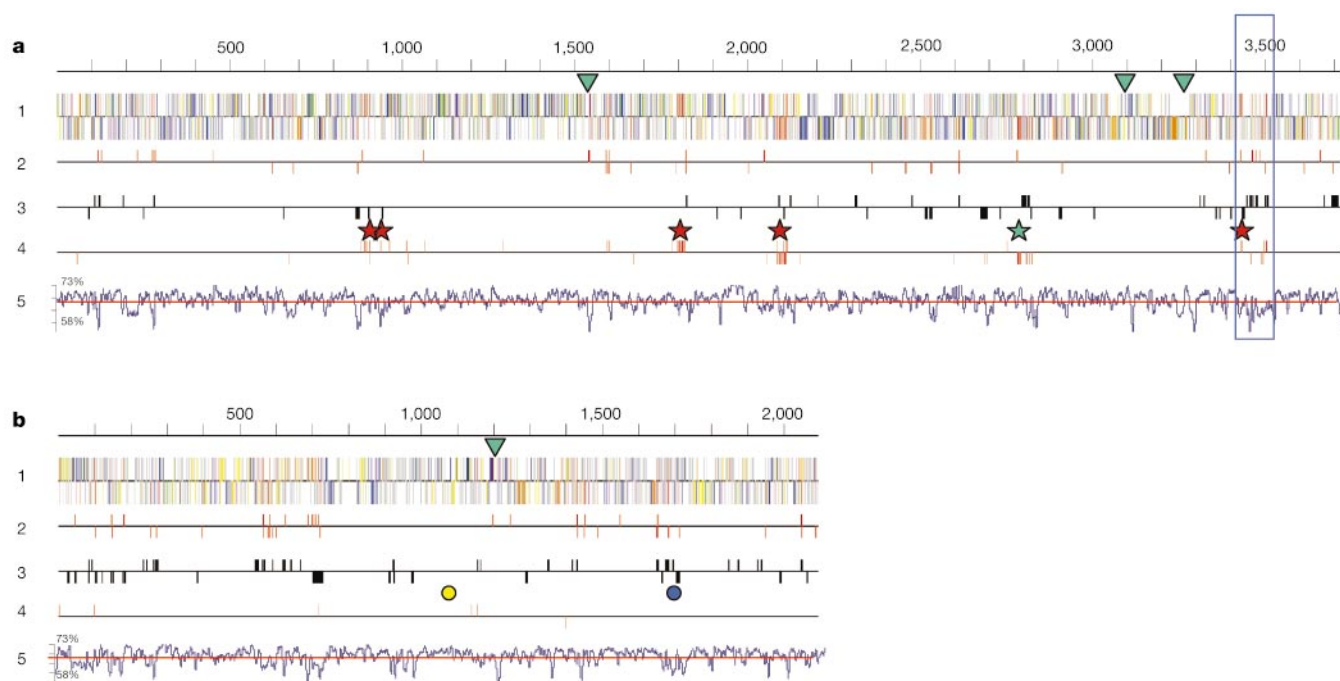


Figure 1 General organization of the *R. solanacearum* strain GMI1000 genome. The two circular replicons are represented linearly starting from base number 1 (**a**, chromosome; **b**, megaplasmid). Kilobases are indicated along the top. For each replicon the distribution of protein-coding genes (line 1), insertion sequences (line 2), ORFs from other genetically mobile elements (line 4) and ACURs (line 3) are represented. The percentage of G+C variation (from the average; red line) along the length of the genome using a 2,500-bp window is shown (line 5). Colours on line 1 correspond to the main classes of genes (the

colour code is available at <http://sequence.toulouse.inra.fr/R.solanacearum.html>). Green triangles point towards the positions of rDNA loci; red stars locate the major bacteriophage remnants; the green star corresponds to the position of the conjugative transposon. The yellow circle indicates the position of the *hrp* locus and the blue circle indicates the tandem duplication of a 31-kb region. The blue box in **a** delineates the region of the genome enlarged in Fig. 2.

haemolysin-like genes belonging to the RTX toxin family²¹ and several peptide or polyketide synthase genes. In particular, the two largest open reading frames (ORFs) in the genome (RS05859 and RS05860, coding for 5,953 and 6,889 amino-acid products, respectively) are highly related to the syringomycin synthase gene, which is required for the production of a *Pseudomonas syringae* toxin²².

An unusual characteristic of the *R. solanacearum* genome is that it contains a large number of genes coding for outer-membrane

proteins or components of bacterial appendages (pili, fimbriae) implicated in the attachment of the bacterium to external surfaces. We found at least 35 genes, distributed in 5 gene clusters, involved in the biogenesis of a type IV pilus. Type IV pili are known from several other bacterial systems to be adhesion factors, and are responsible for movement of bacteria over epithelial surfaces without the use of flagella²³. Furthermore, two other gene clusters are predicted to encode an unusual type of pilus structure that mediates a

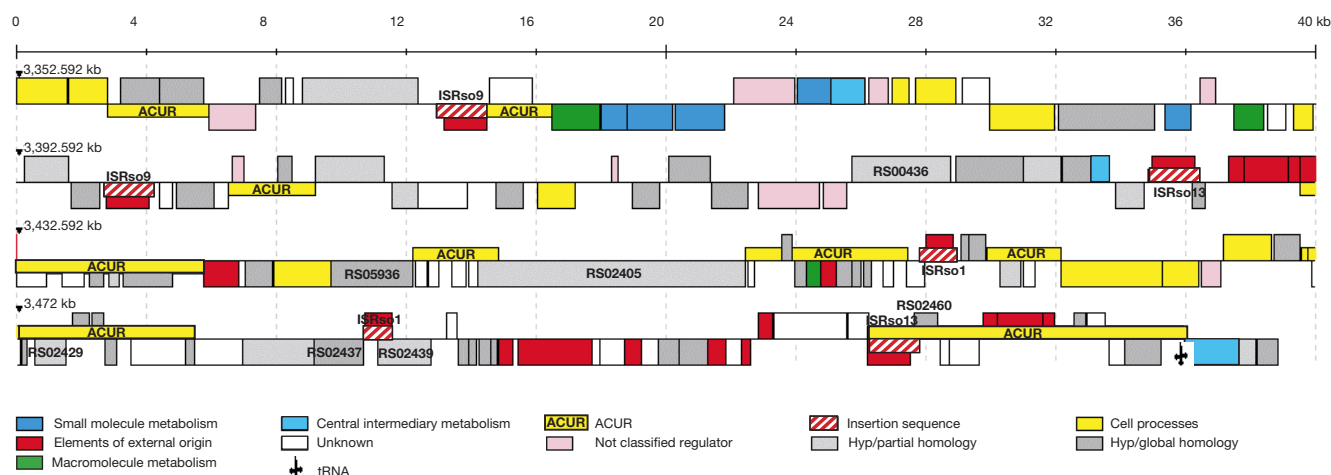


Figure 2 Enlarged view of part of the chromosome. This figure reveals the mosaic structure of the genome. ACURs (narrow yellow rectangles) alternate with genes from mobile genetic elements (large red rectangles with insertion sequences shown in striped white and red rectangles) and with genes that fit the standard Markov model for *R. solanacearum* genes. An Ala-tRNA gene is at the end of the last ACUR. This region

contains genes encoding two candidate type-III secreted effectors (RS02429 and RS02460), three haemagglutinin-related proteins (RS00436, RS02405 and RS5936) and a Vgr-related protein (RS02437 and RS02439) inactivated by insertion of ISRs01. Colour codes for other genes are given at the bottom of the figure and correspond to the main classes defined in the nomenclature of ref. 46.

Table 2 Known and candidate genes responsible for pathogenesis in the genome of *R. solanacearum* strain GMI1000

Pathogenicity genes	No. of genes
Known	
Type III secretion system and secreted effectors	31
Global regulatory functions	11
Exopolysaccharide biosynthesis	18
Hydrolytic enzymes	4
Hormone production	1
Candidate	
Type III secretion-dependent effectors	51
Type III effectors based on homology	25
Type III effectors based on structural features	26
Adhesion/surface proteins	93
Haemagglutinin-related proteins	27
Type IV fimbrial biogenesis proteins	35
Other pili/fimbriae	24
Hydrolytic enzymes/host cell wall degradation	5
Toxins	13
Resistance to oxidative stress	10
Plant hormones and signalling molecules	7
Others	16

tight adherence to surfaces similar to those reported recently in *Caulobacter crescentus*²⁴ and *Actinobacillus actinomycetemcomitans*²⁵. For all of these pili/fimbriae-coding systems, we found multiple copies of the pilin structural genes. This raises the possibility that these genes are expressed in different contexts or may have slightly different structural roles, thereby broadening the adaptative ability of this wide-host-range pathogen to interact with diverse environmental substrates, including host epidermal surfaces.

Another peculiarity concerning the abundance of adhesion/attachment functions in *R. solanacearum* is exemplified by a class of surface molecules encoded by long ORFs (9 frames with a coding potential greater than 2,500 amino acids). These translated products share homology with proteins that are adhesins in other bacterial pathogens, the filamentous haemagglutinin (FhaB) of *Bordetella pertussis* and the HMW1A/HMW2A adhesins of *Haemophilus influenzae*²⁶. In total, there are 14 probable haemagglutinin-type proteins, and 13 additional ORFs coding for proteins containing variable internal repeats that are structurally related to filamentous haemagglutinins. *Ralstonia solanacearum* therefore has the greatest number of these haemagglutinin-related proteins of all the completed bacterial genomes.

We have also identified a family of related proteins presenting some degree of similarity to the *Agrobacterium tumefaciens* proteins AttM and AttZ, both of which are required for attachment to plant cells and for virulence²⁷. The *R. solanacearum* genome is thus rich in attachment factors, perhaps functioning as determinants for a wide host range.

Effectors dependent on a type III secretion system

Ralstonia solanacearum possesses a cluster of *hrp* genes encoding a type III secretion system (TTSS) that is essential for pathogenicity. Bacterial TTSSs are conserved among both plant and animal pathogens and translocate effector proteins into the cytoplasm of the host cell^{4,28}. Identification of translocated effectors and establishment of their mode of action after delivery into host cells are two chief challenges, the solution of which has high potential for the conception of new therapeutic strategies.

A principal outcome of the analysis of the *R. solanacearum* genome is the discovery of multiple genes related to the avirulence (*avr*) genes, described in several plant pathogenic bacteria. These *avr* genes encode effector proteins presumably injected into host cells through the *hrp*-encoded TTSS²⁸. The 14 ORFs coding for products related to Avr determinants (8 with global homology and 6 with homology restricted to a specific domain) are distributed on both replicons. This finding is notable because to date *avr* genes

have been found by functional assays only in phytopathogenic bacteria with a limited host range (often being confined to members of a single plant species or genus), the host range being molecularly governed by *avr* genes²⁹. Furthermore, the presence of Avr-related proteins in *R. solanacearum* is surprising as no report of *avr*-dependent monogenic resistance of solanaceous crops towards bacterial wilt has been reported. It is probable that these Avr proteins must confer selective advantages, acting collectively as pathogenicity factors on a large set of host plants³⁰. It is also possible that *R. solanacearum* possesses some general suppressor(s) of the plant defence response triggered by the recognition of Avr proteins by plant factors.

A second set of 9 proteins dispersed throughout the genome and potentially transiting through the TTSS was identified by homology to determinants in other plant pathogenic bacteria, located in the immediate regions flanking *hrp* loci. The *hrp* locus of *Pseudomonas syringae*, for example, is part of a pathogenicity island (PAI)³¹ that harbours several Hrp-dependent effector proteins³². A PIP-box consensus motif (TTCGC-N15-TTCGC)³³, suggestive of *hrp*-dependent regulation, is present in the promoter sequence of 6 *avr* gene homologues (RS02644, RS04524, RS05218, RS05356, RS05373 and RS05468).

As the TTSS-dependent pathogenicity factors are thought to be injected directly into eukaryotic cells to exert their anti-host function, we looked for proteins exhibiting typical eukaryotic features and functions. As a result, we identified potential functional homologues of essential pathogenicity effectors found in bacterial pathogens of animals, such as protein kinases (RS01445 and RS05210) and a probable tyrosine phosphatase (RS03075). Another ORF product (RS04706) contains 3.5 internal repeats structurally related to the PPR motif³⁴—a motif prevalent in plant organellar proteins, but never found to date in prokaryotic proteins. Three gene families encode proteins with ankyrin repeat domains³⁵, a pirin-like domain³⁶ and leucine-rich repeats (LRRs)³⁷. Although these classes of proteins are not strictly restricted to eukaryotes, they appear to be implicated in eukaryotic signalling pathways, presumably through the properties of these specific domains to establish protein–protein interactions. Several LRR-containing proteins have been shown to transit through TTSSs in pathogens such as *Yersinia*, *Shigella* or *Salmonella* sp., and recently in *R. solanacearum*^{28,38}. We found a total of 10 genes coding for 3 families of LRR-containing proteins scattered in the genome, drawing attention to the probable functional redundancy of these candidate effectors of pathogenicity. Considering the above observations, we estimate that the number of genes encoding potential TTSS-dependent effectors is 40 or higher. This number is greater than that (25) estimated for the bacterial agent of dysentery *Shigella flexneri*³⁹. Genome-scale studies carried out on other plant-pathogenic bacteria will soon reveal whether this high number of effectors is correlated with the wide host range of *R. solanacearum*.

Evolution of virulence

Although most of the candidate genes encoding TTSS-dependent effectors are not found near the *hrp* gene cluster and are distributed evenly on both replicons, approximately half of them appear to reside within clusters of ACURs. Five of these ACUR clusters have the typical features of PAIs such as the presence of DNA sequences indicative of gene mobility (insertion sequences, transposases) or recombination events (Rhs and Vgr elements), and, in some cases, the association of these regions with tRNA or prophage sequences³¹ (Fig. 2). Moreover, 15 of the approximately 40 candidate TTSS-dependent effectors, identified on the basis of homologies or structural features, have a significantly different G+C content than the mean of the *R. solanacearum* genome (see Supplementary Information Table 2). This suggests that they may have been acquired through horizontal gene transfer. In addition, some observations suggest that ACURs could also contribute to the

evolution of new virulence genes. For example, there are three genes related to the host specificity determinant *pthG* from *Erwinia herbicola*⁴⁰ (RS04524, RS05166 and RS05218), located in three different ACURs. These gene products all have a common PthG amino terminus (54% identical residues over the 35 N-terminal amino acids) but with a variable carboxy terminus. Gene duplication and evolution may possibly account for the generation of new virulence specificities in the C termini while maintaining appropriate expression and targeting signals in the N termini. A similar mechanism of duplication followed by differential evolution of domains can also be observed with members of the haemagglutinin-related protein family (such as RS02101 and RS02405) located within or at the border of ACURs. In addition, the deletion of genes could also serve as a means of bacterial adaptation. Evolutionary models of interactions between plants and plant-pathogenic bacteria state that the emergence of a plant resistance gene that recognizes a virulence gene would abolish the value of the corresponding virulence gene⁴¹. Possible evidence for this hypothesis may be found in the occurrence of single or multiple frameshifts or insertion sequence insertions in putative virulence gene homologues (RS00660, RS03919 and RS02460). These deactivating mechanisms may be the result of an evolutionary elimination of genes that have become liabilities for the pathogen.

A close investigation of each of the 30-kb regions flanking the *hrp* locus did not reveal significant changes in the G+C content nor the presence of DNA mobility associated elements typically observed in ACURs. This indicates that, contrary to the *P. syringae* *hrp* gene cluster and flanking regions³², the *R. solanacearum* *hrp* locus and flanking regions containing virulence genes is not, in the strictest sense, a PAI. Instead, our analysis suggests that this region is composed of a core group of ancestral pathogenicity genes that have been subjected to long coevolution with the *R. solanacearum* genome. Thus, the evolutionary status of the *hrp* region is in sharp contrast with the set of candidate effector-encoding genes that are scattered in the genome and often associated with ACURs. This leads us to speculate that *R. solanacearum*, along with an ancestral TTSS and associated effectors, acquired new effector genes through horizontal transfer, thereby remaining a successful pathogen.

Comparative genomics

A comparison of the *R. solanacearum* proteome with the proteome of bacteria that have been sequenced entirely reveals a close similarity of *R. solanacearum* with two other large-genome bacteria (Supplementary Information Figs 5 and 6), *P. aeruginosa*⁴² and *Sinorhizobium meliloti*⁴³. Like *R. solanacearum*, these two species interact with eukaryotic hosts and can be free living in soil. In particular, the proteome of *R. solanacearum* shares many common features with that of the opportunistic pathogen *P. aeruginosa*: (1) a diversity of nutritional pathways; (2) many membrane transport systems; (3) complex chemosensory systems; and (4) a large number of regulatory genes, with a high proportion of two-component regulatory system proteins (37 sensors, 55 response regulators and 6 sensor-response regulator hybrids in *R. solanacearum*). These features are consistent with the evolution of adaptive responses to changes in environment, permitting the bacteria to thrive in diverse ecological niches.

We used a similar comparative analysis on *R. solanacearum* candidate pathogenicity determinants (Supplementary Information Table 3). Only homologies restricted to degradative enzymes, resistance to oxidative stress, haemolysins, peptide synthases and certain attachment structures were found with the plant pathogen *X. fastidiosa*⁸. The same functions also seem to be widely conserved among other bacterial proteomes, contrarily to most of the predicted TTSS-dependent effectors. Several TTSS-dependent effectors appear to be conserved in the plant-pathogenic bacteria, as demonstrated by conservation of 13 known effector proteins. However, with two exceptions (AvrXv/YopP-related proteins⁴ and possibly

RS02872), these effectors are not found in bacterial pathogens of animals. Significantly, most of these TTSS-dependent effectors are not found in the genome of *Ralstonia metallidurans*, a non-pathogenic bacterium that is taxonomically the closest related species to *R. solanacearum*. Our results imply that bacterial pathogens of plants and animals, despite extremely high conservation of the apparatus used to secrete effectors, harbour distinct arrays of specialized effectors. According to this hypothesis, TTSS-dependent effectors would have different cellular targets or effects on the respective hosts, resulting in the differential cellular processes observed during infection of animal and plant cells. □

Methods

Sequencing and assembly

We generated about 80,000 sequences from both ends of genomic clones ranging from 1.5 to 100 kb. The steps to obtain the final sequence (shotgun assembly, gap closure and polishing) are described in the Supplementary Information. The validity of the sequence was assessed by comparing the restriction enzyme pattern deduced from the sequence to the experimentally observed restriction pattern obtained by digestion of clone DNA of a minimal tilling path composed mainly of bacterial artificial chromosome (BAC) clones with 6-base recognition enzymes. A total of 99.25% of the sequences were validated with at least two different restriction enzymes, 0.49% with only one enzyme, and 0.26% of the sequence was not validated, potentially owing to local mis-assemblies (the largest region is less than 5.1 kb). The location of these regions can be found in the Supplementary Information.

Gene prediction and annotation

Sequence analysis and annotation were performed using iANT (Integrated Annotation Tool)⁴⁴ as described for *S. meliloti*⁴⁵, except that the probabilistic Markov model for coding regions used by the gene prediction software FrameD was constructed on 77 *R. solanacearum* gene sequences obtained from public databanks. The alternative matrix was built using genes first identified in ACURs based on homology, as revealed by BLASTX analysis. Predicted ORFs were reviewed individually by gene annotators for start codon assignment. Output of Prosite search and BLASTP analysis on the corresponding products were also individually expertized to generate the proposed annotations. Proteins were classified according to Riley's rules⁴⁶. The complete annotated genetic map, search tools (SRS, BLAST), annotation and process classification are available at <http://sequence-toulouse.inra.fr/R.solanacearum.html>.

Statistical analysis

We performed statistical analysis using Fisher's exact test⁴⁷.

Received 10 July; accepted 30 November 2001.

- Smith, E. F. A bacterial disease of tomato, pepper, eggplant and Irish potato (*Bacillus solanacearum* nov. sp.). *US Dep. Agric. Div. Vegetable Physiol. Pathol. Bull.* **12**, 1–28 (1896).
- Yabuuchi, E., Kosako, Y., Yano, I., Hotta, H. & Nishiuchi, Y. Transfer of two *Burkholderia* and an *Alcaligenes* species to *Ralstonia* gen. nov.: proposal of *Ralstonia pickettii* (Ralston, Palleroni and Doudoroff 1973) comb. nov., *Ralstonia solanacearum* (Smith 1896) comb. nov. and *Ralstonia eutropha* (Davis 1969) comb. nov. *Microbiol. Immunol.* **39**, 897–904 (1995).
- Hayward, A. C. In *Encyclopedia of Microbiology* Vol. 4 (ed. Lederberg, J.) 32–42 (Academic, San Diego, 2000).
- Staskawicz, B. J., Mudgett, M. B., Dangl, J. L. & Galan, J. E. Common and contrasting themes of plant and animal diseases. *Science* **292**, 2285–2289 (2001).
- Granada, G. A. & Sequeira, L. Survival of *Pseudomonas solanacearum* in soil, rhizosphere and plant roots. *Can. J. Microbiol.* **29**, 433–440 (1983).
- Parkhill, J. et al. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**, 502–505 (2000).
- Tettelin, H. et al. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* **287**, 1809–1815 (2000).
- Simpson, A. J. et al. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406**, 151–157 (2000).
- The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
- Boucher, C., Martinel, A., Barberis, P., Alloing, G. & Zischek, C. Virulence genes are carried by a megaplasmid of the plant pathogen *Pseudomonas solanacearum*. *Mol. Gen. Genet.* **205**, 270–275 (1986).
- Rosenberg, C., Casse-Delbart, F., Dusha, I., David, M. & Boucher, C. Megaplasmids in the plant associated bacteria *Rhizobium meliloti* and *Pseudomonas solanacearum*. *J. Bacteriol.* **150**, 402–406 (1982).
- Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13**, 660–665 (1996).
- Boucher, C., Barberis, P. & Arlat, M. Acridine orange selects for deletion of *hrp* genes in all races of *Pseudomonas solanacearum*. *Mol. Plant-Microbe Interact.* **1**, 282–288 (1988).
- Bertolla, F., Van Gijsegem, F., Nesme, X. & Simonet, P. Conditions for natural transformation of *Ralstonia solanacearum*. *Appl. Environ. Microbiol.* **63**, 4965–4968 (1997).
- Oggioni, M. R. & Claverys, J. P. Repeated extragenic sequences in prokaryotic genomes: a proposal for the origin and dynamics of the RUP element in *Streptococcus pneumoniae*. *Microbiology* **145**, 2647–2653 (1999).

16. Merlin, C., Springael, D., Mergeay, M. & Toussaint, A. Organisation of the *bph* gene cluster of transposon Tn4371, encoding enzymes for the degradation of biphenyl and 4-chlorobiphenyl compounds. *Mol. Gen. Genet.* **253**, 499–506 (1997).
17. Wang, Y. D., Zhao, S. & Hill, C. W. Rhs elements comprise three subfamilies which diverged prior to acquisition by *Escherichia coli*. *J. Bacteriol.* **180**, 4102–4110 (1998).
18. Wilderman, P. J., Vasil, A. I., Johnson, Z. & Vasil, M. L. Genetic and biochemical analyses of eukaryotic-like phospholipase D of *Pseudomonas aeruginosa* suggest horizontal acquisition and a role for persistence in a chronic pulmonary infection model. *Mol. Microbiol.* **39**, 291–303 (2001).
19. Brumbley, S. M., Carney, B. F. & Denny, T. P. Phenotype conversion in *Pseudomonas solanacearum* due to spontaneous inactivation of PhcA, a putative LysR transcriptional regulator. *J. Bacteriol.* **175**, 5477–5487 (1993).
20. Alvarez, M. E. *et al.* Reactive oxygen intermediates mediate a systemic signal network in the establishment of plant immunity. *Cell* **92**, 773–784 (1998).
21. Lally, E. T., Blake Hill, R., Kieba, I. R. & Korostoff, J. The interaction between RTX toxins and target cells. *Trends Microbiol.* **7**, 356–361 (1999).
22. Bender, C. L., Alarcon-Chaidez, F. & Gross, D. C. *Pseudomonas syringae* phytotoxins: mode of action, regulation, biosynthesis by peptide and polyketide synthases. *Microbiol. Mol. Biol. Rev.* **63**, 266–292 (1999).
23. Wall, D. & Kaiser, D. Type IV pili and cell motility. *Mol. Microbiol.* **32**, 1–10 (1999).
24. Skerker, J. M. & Shapiro, L. Identification and cell cycle control of a novel pilus system in *Caulobacter crescentus*. *EMBO J.* **19**, 3223–3234 (2000).
25. Kachlany, S. C. *et al.* *flp-1*, the first representative of a new pilin gene subfamily, is required for non-specific adherence of *Actinobacillus actinomycetemcomitans*. *Mol. Microbiol.* **40**, 542–554 (2001).
26. Jacob-Dubuisson, F., Locht, C. & Antoine, R. Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. *Mol. Microbiol.* **40**, 306–313 (2001).
27. Matthisse, A. G., Yarnall, H., Boles, S. B. & McMahan, S. A region of the *Agrobacterium tumefaciens* chromosome containing genes required for virulence and attachment to host cells. *Biochim. Biophys. Acta* **1490**, 208–212 (2000).
28. Cornelis, G. R. & Van Gijsegem, F. Assembly and function of type III secretory systems. *Annu. Rev. Microbiol.* **54**, 735–774 (2000).
29. Leach, J. E. & White, F. F. Bacterial avirulence genes. *Annu. Rev. Phytopathol.* **34**, 153–179 (1996).
30. Kjemtrup, S., Nimchuk, Z. & Dangel, J. L. Effector proteins of phytopathogenic bacteria: bifunctional signals in virulence and host recognition. *Curr. Opin. Microbiol.* **3**, 73–78 (2000).
31. Hacker, J. & Kaper, J. B. Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* **54**, 641–679 (2000).
32. Alfano, J. R. *et al.* The *Pseudomonas syringae* Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. *Proc. Natl Acad. Sci. USA* **97**, 4856–4861 (2000).
33. Fenselau, S. & Bonas, U. Sequence and expression analysis of the *hrpB* pathogenicity operon of *Xanthomonas campestris* pv. *vesicatoria* which encodes eight proteins with similarity to components of the Hrp, Ysc, Spa, and Fli secretion systems. *Mol. Plant-Microbe Interact.* **8**, 845–854 (1995).
34. Small, I. D. & Peters, N. The PPR motif—a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* **25**, 46–47 (2000).
35. Bork, P. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins* **17**, 363–374 (1993).
36. Wendler, W. M., Kremmer, E., Forster, R. & Winnacker, E. L. Identification of pirin, a novel highly conserved nuclear protein. *J. Biol. Chem.* **272**, 8482–8489 (1997).
37. Kajava, A. V. Structural diversity of leucine-rich repeat proteins. *J. Mol. Biol.* **277**, 519–527 (1998).
38. Guéron, M., Timmers, A. C., Boucher, C. & Arlat, M. Two novel proteins, PopB, which has functional nuclear localization signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of *Ralstonia solanacearum*. *Mol. Microbiol.* **36**, 261–277 (2000).
39. Buchreiser, C. *et al.* The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol. Microbiol.* **38**, 760–771 (2000).
40. Ezra, D., Barash, I., Valinsky, L. & Manulis, S. The dual function in virulence and host range restriction of a gene isolated from the pPATH (Ehg) plasmid of *Erwinia herbicola* pv. *gypsophillae*. *Mol. Plant-Microbe Interact.* **13**, 683–692 (2000).
41. Alfano, J. R. & Collmer, A. Bacterial pathogens in plants: life against the wall. *Plant Cell* **8**, 1683–1698 (1996).
42. Stover, C. K. *et al.* Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an opportunistic pathogen. *Nature* **406**, 959–964 (2000).
43. Galibert, F. *et al.* The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* **293**, 668–672 (2001).
44. Thébault, P., Servant, F., Schiex, T., Kahn, D. & Gouzy, J. in *JOBIM Conf. Proc.* 361–365 (ENSA and LIRM Editor, Montpellier, 2000).
45. Capela, D. *et al.* Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proc. Natl Acad. Sci. USA* **98**, 9877–9882 (2001).
46. Karp, P. D. *et al.* EcoCyc: Encyclopedia for *Escherichia coli* genes and metabolism. *Nucleic Acids Res.* **27**, 55–58 (1999).
47. Fischer, R. A. On the interpretation of chi-square from contingency tables, and the calculation of P. *J. R. Stat. Soc.* **85**, 87–94 (1922).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank N. Aiache and C. Cruaud for technical assistance. The Laboratoire de Biologie Moléculaire des Interactions Plantes Microorganismes is supported by INRA, CNRS and Toulouse Genopole. M. Whalen was supported by the National Institute for Health.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.A.B. (e-mail: boucher@toulouse.inra.fr). The sequences for the chromosome and megaplasmid are deposited in the EMBL database under accession numbers AL646052 and AL646053, respectively.

Topologically protected quantum bits using Josephson junction arrays

L. B. Ioffe^{*†}, M. V. Feigel'man[†], A. Ioselevich[†], D. Ivanov[‡], M. Troyer[‡] & G. Blatter[‡]

^{*} Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA

[†] Landau Institute for Theoretical Physics, 117940 Moscow, Russia

[‡] Theoretische Physik, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

All physical implementations of quantum bits (or qubits, the logical elements in a putative quantum computer) must overcome conflicting requirements: the qubits should be manipulable through external signals, while remaining isolated from their environment. Proposals based on quantum optics emphasize optimal isolation^{1–3}, while those following the solid-state route exploit the variability and scalability of nanoscale fabrication techniques^{4–8}. Recently, various designs using superconducting structures have been successfully tested for quantum coherent operation^{9–11}, however, the ultimate goal of reaching coherent evolution over thousands of elementary operations remains a formidable task. Protecting qubits from decoherence by exploiting topological stability is a qualitatively new proposal¹² that holds promise for long decoherence times, but its physical implementation has remained unclear. Here we show how strongly correlated systems developing an isolated twofold degenerate quantum dimer liquid ground state can be used in the construction of topologically stable qubits; we discuss their implementation using Josephson junction arrays. Although the complexity of their architecture challenges the technology base available today, such topological qubits greatly benefit from their built-in fault-tolerance.

Any quantum computer has to incorporate some fault tolerance because we cannot hope to eliminate every source of decoherence. Quantum error-correction schemes¹³ have been developed using redundant multi-qubit encoding of the quantum data combined with error-detection and recovery steps. Such error-correction schemes are generic (and hence are applicable to any hardware implementation), but require repeated active interference with the computer during run-time; the delocalization of the data, often in a hierarchical structure, boosts the system size by a factor of 10^2 to 10^3 . Delocalization of the quantum information is also at the heart of topological quantum computing¹²; however, the stabilization against decoherence is deferred to the hardware level (and so it is tied to the specific implementation) and is achieved passively. In searching for a physical implementation of topological qubits we strive for an extended (many body) quantum system where the Hilbert space of quantum states decomposes into mutually orthogonal sectors, each sector remaining isolated under the action of local perturbations. Choosing the two qubit states from ground states in different sectors protects these states from unwanted mixing through noise; protection from leakage within the sector has to be secured through a gapped excitation spectrum. As no local operator can interfere with these states, global operators must be found (and implemented) that allow the manipulation of the qubit state.

A promising candidate that fulfils the above requirements is the quantum dimer system^{14–17}. Its simplest version is defined on a square lattice and has been discussed in the context of short-range resonating-valence-bond (RVB) models¹⁸ for high-temperature superconductivity¹⁴. In order to exploit its topological properties for quantum computing we use an implementation on the triangular lattice: allowed configurations are coverings with dimers connecting neighbouring vertices, satisfying the constraint that

every vertex belongs to only one dimer (see Fig. 1). Attributing equal energies to these states results in an extensive (classical) degeneracy; the latter is lifted in the quantum version with the hamiltonian $H_{\text{dim}} = H_{\square} + H_{\triangle} + H_{\diamond}$, where

$$H_{\square} = \sum_{\square} [-t(|-\rangle\langle-/| + |/\rangle\langle-|) + v(|-\rangle\langle-| + |/\rangle\langle-/|)]$$

and corresponding definitions for $H_{\triangle}$ and $H_{\diamond}$. The kinetic term (proportional to t) rotates parallel dimers (shown as parallel symbols above) on appropriate plaquettes $\square$, $\triangle$ and $\diamond$ and pushes the 'columnar' and 'staggered' phases (see Fig. 1) to minimal and maximal energies, respectively. The interaction term (proportional to v) allows frustration of parallel dimers, counteracting these orderings and producing a dimer-liquid phase. Whereas the order characterizing the crystal phases is obvious, the topological order present in the dimer liquid is more subtle^{14–16}: consider a cylindrical geometry with periodic boundary conditions along the x axis, leaving the boundaries along the y direction free (see Fig. 1). The parity of the dimer count along the reference line γ parallel to the y axis defines a topological order parameter: the action of the hamiltonian H_{dim} leaves the parity invariant and the Hilbert space splits into two topological sectors $\mathcal{H}_e$ and $\mathcal{H}_o$ characterized by even and odd dimer counts.

The particular features of this model we wish to exploit for quantum computing are (1) the presence of a liquid ground state with its robustness to disorder (see below), and (2) the topological structure of the Hilbert space. We note that the liquid phase of the square lattice dimer model is restricted to the point $t = v$ (ref. 15) and its Hilbert space decays into infinitely many sectors¹⁷; these problematic features are overcome in the triangular lattice. Indeed, evidence for a dimer liquid phase extending over a finite interval in the parameter v/t has been provided by numerical simulations: recent Monte Carlo simulations¹⁹ on large systems with $L_{x,y} = 36$ exhibit short-range dimer–dimer correlations within a parameter region $2/3 < v/t < 1$, indicative of a liquid state (see refs 20 and 21 for a discussion of similar exotic groundstates in spin models). Also, the weak temperature dependence of this data within the temperature interval $0.25t > T > 0.03t$ provides evidence for a gapped spectrum. These results have been corroborated by our numerical diagonalization studies on both cylindrical (periodicity along the x direction) and toroidal geometries with system sizes going up to $L_{x,y} = 6$. We find: (1) the triangular dimer model develops an isolated twofold degenerate dimer liquid ground state for a cylinder (it is fourfold degenerate for the toroidal geometry) within the

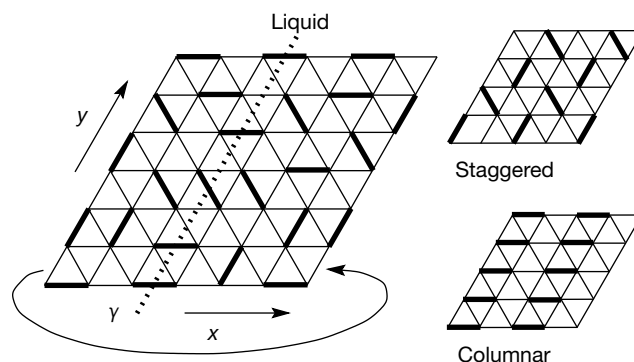


Figure 1 Dimers on a triangular lattice forming liquid (main panel), staggered, and columnar configurations. Periodic boundary conditions are applied along the x direction for the cylindrical geometry and along both x and y directions for the toroidal set-up. The parity of the dimer count along the reference line γ remains unchanged under the action of the dimer hamiltonian H_{dim} and hence defines a topological invariant. Correspondingly, the Hilbert space $\mathcal{H}$ splits into two sectors $\mathcal{H}_e$ and $\mathcal{H}_o$ hosting states with even and odd parity, respectively.

parameter region $\sim 0.8 < v/t < 1$ with a gap Δ of order 0.1t separating the ground state from excited states (exact degeneracy is realized in the thermodynamic limit and for particular geometries). The mixing of the two groundstates involves creation of topological defects (real or virtual) violating the dimer constraint and is exponentially suppressed in the system size L_x . (2) Comparing spectra of samples with cylindrical and toroidal boundary conditions we verify the absence of low-lying edge states. (3) By perturbing the system with a quenched disorder potential we find that the ground-state degeneracy is robust to within a factor of 10^{-3} to 10^{-2} of the disorder potential (see Fig. 2). We expect this robustness to increase exponentially in the system size L_y (see Methods). We note that the amplitude mixing and the robustness against disorder involve different physical mechanisms with separate correlation lengths.

Translating these findings to potential physical implementations of this two-state system, we find the system to be quite tolerant with respect to (static) local variations in parameters appearing in the fabrication process. Regarding decoherence, the amplitude mixing of the two ground states is strongly suppressed as it requires creation of topological defects. The decoherence in the relative phase of the two ground states originates either from their adiabatic splitting by an external low-frequency noise, or the creation of non-topological excitations within each sector. The impact of noise is suppressed for the same reason that leads to the robustness with respect to local disorder, while excitations are inhibited by the gap Δ , which sets the requirement $T \ll \Delta$ on the operation temperature.

We propose two types of Josephson junction arrays for the implementation of the above quantum dimer liquid ground states and their use for quantum computing. In the first array (JJT), the

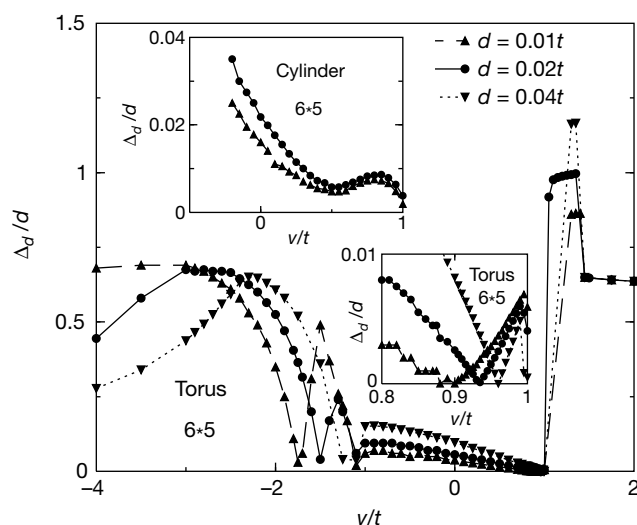


Figure 2 Splitting Δ_d of the ground-state energies under the action of a disorder potential of strength d for a 6×5 torus and a 6×5 cylinder (top inset; periodicity is along x with $L_x = 6$) testing the susceptibility of the degenerate ground states to local perturbations. Random chemical potentials (homogeneously distributed over the interval $[-d/2, d/2]$ with $d < t$) are assigned to the links. The disorder-induced ground-state splitting Δ_d collapses suddenly at the transition from the crystal ($v > t$) to the liquid phase ($v < t$) and slowly recovers as v is further decreased. We attribute the sharp drop at $v \approx t$ to the first-order solid-liquid transition; the small disorder splitting at $v < t$ indicates the efficient protection of the dimer-liquid ground states from local perturbations (the bottom inset shows an expanded view near $v \approx t$ for the 6×5 torus disclosing a disorder splitting Δ_d of order 10^{-3} to 10^{-2} of the disorder energy d). Decreasing v further below t we observe level crossings which we attribute to the appearance of intermediate crystalline phases (compare with ref. 19) before the fluctuating columnar phase is reached at large negative values of v .

vertices of the triangular lattice are structured into six Y-shaped superconducting islands with two ends joining into the hexagonal vertex and the third end linking to the neighbouring hexagon (see Fig. 3). The capacitances C_Y , C_h , and C_l (of the Y-shaped islands, the hexagons and the links) and the Josephson currents I_h and I_l define the corresponding energies $E^C = e^2/2C$ and $E^J = \Phi_0 I/2\pi c$ and are chosen to emulate the triangular quantum dimer model (for example, $E_h^C = (2e)^2/2C_h$, $E_l^C = \Phi_0 I_h/2\pi c$; $\Phi_0 = hc/2e$ denotes the flux quantum with h , c , and e denoting Planck's constant, the velocity of light, and the elementary charge, respectively). We first find C_Y , C_h and I_l defining the classical dimer states. We choose a large capacitance C_h in order to join the islands electrically into one hexagonal vertex. A small capacitance C_Y defines the large charging energy $E_{\text{hex}} \approx E_Y^C/6$ of the hexagonal vertex, the basic energy scale of the array. We introduce 'charge frustration' of the vertices by biasing the array with a global electric gate to equalize the energies (to an accuracy better than E_l^J) of two states differing by one Cooper pair (this defines a 'half-filled' array); the large charging energy E_{hex} lifts other charge states to high energies. The Cooper pairs lower their energy via tunnelling (involving the coupling I_l) through the link junction joining two vertices—the 'bonding' state of such a Cooper pair defines the dimer state. With only half a Cooper pair available per hexagon, each vertex is involved in the formation of only one of these 'valence bonds'. This defines the 'classical' configurations of the dimer model; the corresponding space of dimer states is protected by the energy scale E_l^J .

The dimer dynamics involves the vertex junction (with weak coupling $I_h \ll I_l$) and proceeds via localization of one dimer to the vertex (with an energy cost E_l^J) and subsequent hops (of two parallel dimers or Cooper pairs over junctions with energy E_l^J ; see Fig. 3) to the new vertex islands. The hopping amplitude is of order $t \approx E_h^{J2}/E_l^J$; a perturbative calculation gives $t = (9/16)E_h^{J2}/E_l^J$. The electrostatic interaction between dimers depends on the choice of the capacitance matrix and its dependence on the dimer config-

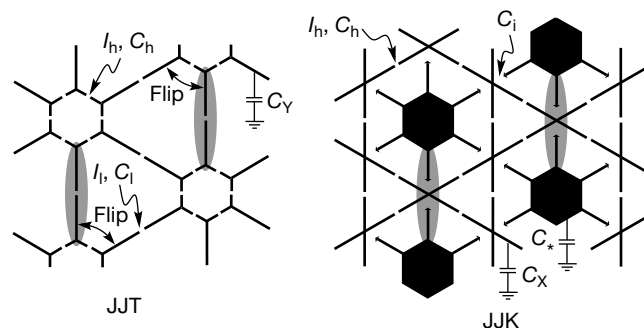


Figure 3 Josephson junction arrays forming triangular (JJT) and Kagome (JJK) lattices and emulating dimer systems. The JJT lattice involves the following parameters: the capacitance C_Y of the island to the groundplate, the capacitance C_h and Josephson current I_h of the hexagonal vertex junction, and the capacitance C_l and Josephson current I_l of the link junction. The large capacitance C_h joins hexagonal islands into electric units which are frustrated by a global gate to accept half a Cooper pair. Cooper pairs resonating between hexagonal islands form dimers (shaded ellipses); dimer flips due to I_h are indicated by arrows. In the JJK lattice, Cooper pairs on X-shaped islands are capacitively coupled through star-shaped islands (C_X and C_* denote the capacitances of the islands to the ground). The gate is tuned to allow for half a Cooper pair per hexagon. Dimers are pairs residing on X-shaped islands and polarizing adjacent star-shaped islands. The similarity between the Kagome lattice and the triangular array becomes obvious when joining pairs of Y-shaped islands (corresponding to the limit $C_l \rightarrow \infty$ and $I_l \rightarrow \infty$) and contracting the link to obtain the X-shaped island—the vertex junctions in the JJT array now play the role of the junctions on the Kagome lattice, hence the same index h is chosen for the junctions on the hexagons; note that the JJK array involves only one type of Josephson junction.

uration is non-trivial. First, we compare the energies of the staggered and columnar configurations and find the interaction energy ν between parallel dimers; in the limit $C_1 \ll C_Y$, C_h , $\nu = E_h^C(C_1/C_h)[(1 + C_Y/C_h)(1 + 3C_Y/C_h)]^{-2}$. Second, we have checked that the electrostatic energies of the liquid-phase dimer configurations indeed scale (to within $\sim 5\%$ accuracy) with the number of parallel dimer pairs. The condition $C_1 \ll C_Y$ guarantees a short-range interaction between dimers.

The backbone of the second array (JJK) is made from X-shaped islands arranged into a Kagome lattice with capacitive and Josephson couplings E_h^C and E_h^J (see Fig. 3). A second triangular lattice of star-shaped islands introduced into the hexagons of the Kagome lattice is only capacitively coupled (with energy E_i^C) to the islands of the Kagome lattice. The Kagome sites are biased to accept half a Cooper pair per hexagon. By analogy, dimers now correspond to Cooper pairs localized onto the X-islands of the Kagome lattice. As dimers should not touch one another, no two bosons are allowed on the same hexagon. We choose small capacitances C_h in order to isolate hexagons from one another (note the difference to the JJT array) and large capacitances C_i to join the six X-shaped islands forming the hexagon into one electrostatic unit via their strong coupling to the central star-shaped island. The capacitances C_X , C_* , and C_i then define the charging energy required to put two Cooper pairs on the same hexagon, $E_{\text{hex}} \approx (C_i/C_X)^2 E_*$, which is our basic energy scale (we assume $C_i < C_*$, C_X in order to guarantee good screening on large distances).

The motion of the Cooper pairs (the dimer dynamics) involves the hopping amplitude E_h^J and the charging energy E_{hex} of the virtual state with two Cooper pairs on the same hexagon, $t \approx E_h^J/E_{\text{hex}}$; note that we require $E_h^J \ll E_{\text{hex}}$ not to perturb the proper frustration of the array. The interaction energy ν again derives from a comparison between the staggered and columnar states and has to meet the conflicting requirements of increasing the energy of parallel dimers (next-nearest-neighbour interaction) while leaving non-parallel dimer pairs (next-next-nearest-neighbour interaction) approximately unperturbed. To produce a positive ν we need a finite coupling C_h , which in turn reduces the basic energy E_{hex} . We find that the set $5C_i = C_* = C_X = 20C_h$ produces a (renormalized) protective energy $E_{\text{hex}}^r \approx 0.008E_* \approx 0.2E_{\text{hex}}$ and a repulsive energy $\nu \approx 0.1E_{\text{hex}}^r$. Choosing $E_h^J \approx 0.3E_{\text{hex}}^r$ we can properly satisfy the condition $t \approx \nu$.

We construct a topologically protected qubit from the two-level system defined by the two ground states of a quantum dimer system with cylindrical boundary conditions. We make use of the above Josephson junction arrays and a ring geometry; the two qubit states $|e\rangle$ and $|o\rangle$ are then distinct through the parity of the dimer count along the line γ joining the inner and outer boundaries of the array, see Fig. 4. In order to manipulate the qubit we have to implement the qubit hamiltonian

$$H_{\text{qubit}} = h_x \sigma_x + h_z \sigma_z$$

where σ_x and σ_z are Pauli matrices and h_x , h_z are the (manipulable) parameters producing the amplitude (α) and phase (χ) mixing in the qubit state $|\alpha, \chi\rangle = [|e\rangle + \alpha \exp(i\chi)|o\rangle]/\sqrt{1 + \alpha^2}$. The implementation of h_x requires a controlled mixing of the protected dimer states, implying a reduction of the ground state's topological protection. In the JJT array this is achieved by breaking one dimer bond and the creation of a virtual particle-hole excitation where one Cooper pair retreats to one hexagon (particle), leaving the partner hexagon empty (hole). This virtual excitation costs the energy E_i^J of that particular bond and takes the system (virtually) out of the protected dimer space. While the particle remains pinned to the weak junction, the hole is taken around the inner boundary through appropriate dimer flips and is subsequently recombined with the particle. This process results in a mixing amplitude $h_x \approx E_i^J(E_h^J/E_i^J)^M$, where M denotes the number of links on the inner boundary (this estimate applies to

the 'optimal' dimer configuration shown in Fig. 4). Hence introducing one switchable link junction near the inner boundary allows us to tune the parameter E_i^J and change the mixing amplitude h_x by many orders of magnitude. The variant for the JJK array involves a virtual excitation with two Cooper pairs on one hexagon.

The 'phase shifter' can be implemented through a gated superconducting strip, capacitively coupled to the array and attracting dimers onto the reference line γ . Upon biasing the strip, the energies of the two groundstates are shifted with respect to one another and the qubit's phase χ is modified accordingly. The implementation requires careful adiabatic switching, with the amplitude u (the energy acting on one dimer) and the duration τ of the manipulation limited by the constraints $u \approx t \ll E_i^J$ and $\tau > \hbar/\Delta$ in order to avoid excitations within the dimer liquid. Furthermore, we have to break up the strip during idle time, as the fully connected strip represents a global operator (otherwise, electric fluctuations fed to the strip would decohere the system). Thus the strip has to be constructed from isolated superconducting islands which are to be connected only during the time where the phase shifter becomes operational. Fast switches accomplishing such a connection can be built from superconducting Cooper pair transistors²².

The qubit's manipulation involves strong modifications of the system: the 'amplitude shifter' requires leaving the protected Hilbert space through virtual breaking of a dimer, while the phase shifter requires introduction of a global operator. In turn, these processes are strongly inhibited during idle time and give the qubit its robustness: amplitude mixing is exponentially small in L_x while phase drifting is exponentially small in L_y .

Constructing a register of qubits is surprisingly simple: the same basic Josephson junction array can accommodate K qubits in a geometry with K holes; see Fig. 4. Individual qubit operations are carried out as described above. The implementation of a non-trivial two-qubit operation again involves a superconducting strip: biasing the strip connecting two qubits lifts the energy of the states $|eo\rangle$ and $|oe\rangle$ with respect to the states $|ee\rangle$ and $|oo\rangle$ and combining this two-qubit 'phase shifter' with a suitable set of single qubit

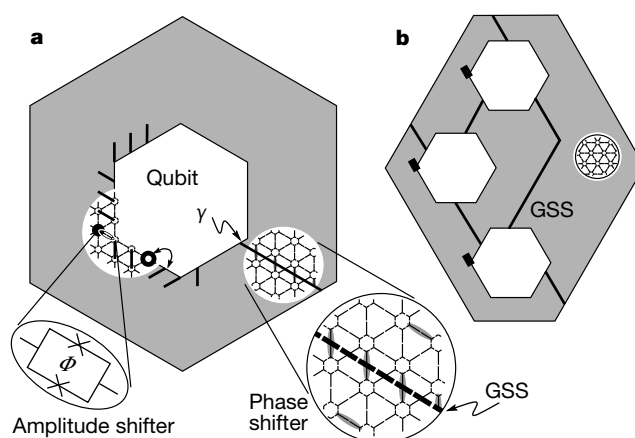


Figure 4 Topologically protected qubits. **a**, The topologically protected qubit involves the two degenerate ground states of the quantum dimer liquid state realized within a triangular Josephson junction array. The qubit's phase is manipulated through the gated superconducting strip (GSS) attracting dimers along γ . The amplitude shifter mixing the ground states is realized with the help of a tunable Josephson junction placed at the inner qubit boundary. Dimer-flips, like a row of falling dominoes, take the hole around the inner boundary. **b**, Array of topologically protected qubits embedded in a Josephson junction array emulating a dimer system. The entanglement of two qubits proceeds through coupling over a gated superconducting strip.

operations allows for the construction of the controlled-NOT operation⁸.

In a first step towards the implementation of topologically protected qubits one may wish to test the proper functionality of the quantum Josephson junction array. Such a test is provided by a measurement of the magnetic susceptibility χ of the ring structure: applying a magnetic field, the mixing energy h_x picks up a factor $\cos(2\pi\Phi/\Phi_0)$, where $\Phi_0 = hc/e$ is the normal state flux quantum. This factor originates from the Aharonov–Bohm phase picked up by the virtual charged excitations encircling the hole. In the absence of disorder the ring's energy changes with flux, $E_{\text{ring}}(\Phi) = h_x |\cos(2\pi\Phi/\Phi_0)|$, resulting in a susceptibility $\chi \propto \partial_\Phi E_{\text{ring}}$ with period $\Phi_0/2$ and a sharp feature at $\Phi = \Phi_0/4$. The (disorder- or finite-size induced) splitting Δ_d in the ground-state energies smears this feature over a region $\delta\Phi/\Phi_0 \approx \Delta_d/h_x$; hence a measurement of $\chi(\Phi)$ allows us to check on the mixing amplitude h_x and on the level splitting.

A particular challenge faced by qubit implementations based on Josephson junctions operating in the charge limit is the presence of stray charges^{5,9}; the robustness in our device efficiently suppresses fluctuations in the gate potential once they drop below t . The problem of stray charges is avoided in designs using the dual phase variable^{7,8,11} and work in this direction is in progress. $\square$

Methods

Additional insight into the physics of the quantum dimer model is gained in a comparison between its square-lattice and triangular versions and their relation to compact quantum electrodynamics. The square lattice dimer model described by the hamiltonian

$$H = \sum_{\square} [-t(|\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|) + v(|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|)]$$

develops a phase diagram similar to the one of the triangular system with columnar ($v < t$), liquid (at the isolated point $v = t$), and brickwall ($v > t$) phases. Its topological structure differs from that of the triangular version: splitting the bipartite lattice into A and B sublattices we attribute dimers a direction (from A to B) and count those pointing from A to B sites (B to A sites) positively (negatively). The topological invariant ϕ counting the weighted number of dimers crossing the reference line γ then takes on all integer numbers and we obtain corresponding topological sectors $\mathcal{H}_\phi$ for each integer ϕ . Adding a term

$$H_d = \sum_{\square} [-t(|\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|) + v(|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|) + \mu(|\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|)]$$

that transforms horizontally ($|\downarrow\downarrow\rangle$) and vertically ($|\uparrow\uparrow\rangle$) shifted dimers into diagonal ones $|\uparrow\downarrow\rangle$ at the energy cost of 2μ (one family of diagonals is chosen) allows us to study the evolution of the dimer model from a square to a triangular lattice via tuning of the chemical potential μ (ref. 24).

The mapping to quantum electrodynamics proceeds along the following steps^{17,23,24}. We decorate the bipartite square lattice with static charges $\rho_{\square} = \pm 1$, placing all positive charges on one sublattice. A dimer covering $\{d_{i,j}\}$ of bonds (i,j) of this 'ionic crystal' is mapped to a distribution of electric dipoles $\{-\mathbf{p}_{i,j}\}$; the electric field $\mathbf{e} = -\langle\mathbf{p}_{i,j}\rangle_{\delta\Lambda}$ resulting from averaging over small areas $\delta\Lambda$ satisfies Gauss' law $\nabla\mathbf{e} = \langle\rho\rangle_{\delta\Lambda} = 0$. The simplest dynamics preserving this constraint is given by Faraday's law or, more precisely, (compact) quantum electrodynamics. Here, we are interested in large-scale (global) aspects and thus keep the discussion on a phenomenological level; the physical properties of the dimer liquid phase will be described in terms of the electric field $\mathbf{e}$ and the dielectric constant ϵ .

Consider first the situation in the absence of diagonal dimers ($\mu \rightarrow \infty$). The topological invariant then is given by the flux $\phi = \oint d\mathbf{y} \epsilon_{xy} \mathbf{e}$ of the electric induction across γ and each sector $\mathcal{H}_\phi$ is characterized by an average polarization $e_x = \phi/L_x$, producing a ground-state energy $E = e^2 L_x L_y / 2 = \phi^2 L_x / 2\epsilon L_y = E_\phi$. A finite value of the chemical potential μ leads to the appearance of mobile defects with charge ± 2 (refs 24, 25). These charges mix states with electric flux $\phi \pm 2$: creating two defects of charge ± 2 , carrying one defect around the cylinder, and subsequent annihilation with its partner changes the flux ϕ by two; as a result, the infinite set of sectors $\mathcal{H}_\phi$ collapses to the even and odd sectors $\mathcal{H}_{\epsilon,\phi}$.

The hamiltonian projected onto the space of ground states within topological sectors $\mathcal{H}_\phi$ takes the form $H_{\text{sec}} = \sum_{\phi} (E_\phi |\phi\rangle\langle\phi| + J(|\phi\rangle\langle\phi+2| + \text{h.c.}))$. For large values of the chemical potential $\mu \gg t$ the defect pairs are only virtual and the amplitude J mixing states with different ϕ is exponentially small in the circumference L_x , $J \propto (t/\mu)^{L_x}$. As μ drops below an energy of order t , defects proliferate and undergo Bose condensation; the concomitant long-range order in the condensate phase φ implies large charge fluctuations and thus the electric field flux across γ fluctuates strongly. In this phase the mixing between sectors is large, with $\langle\phi^2\rangle \propto L_x$; assuming a superfluid density ρ_s , the hydrodynamic

analysis of the charged superfluid provides us with a density–density correlator $\langle\rho_{\mathbf{k},\omega}\rho_{-\mathbf{k},-\omega}\rangle = \rho_s k^2 / (\rho_s \epsilon + \omega^2)$ and the relation $\epsilon \nabla \cdot \mathbf{e} = \rho$ allows us to find the mean squared electric flux crossing the line γ , $\langle\phi^2\rangle = \sqrt{\rho_s \epsilon} L_y / 2$.

The new ground states $|e, \phi\rangle \propto \sum_{\epsilon,\phi} a_{\epsilon,\phi} |\phi_{\epsilon,\phi}\rangle$ involve contributions from all sectors $\mathcal{H}_{\epsilon,\phi}$ with $\phi_\epsilon = 2k$ and $\phi_o = 2k+1$ running over even and odd integers; their physical properties are conveniently characterized by the distribution function $P(\phi) = |a_\phi|^2 \propto \exp(-\phi^2/2\sigma)$, with the gaussian form appropriate for a process involving a sum of statistically independent elements. The width σ is obtained through a comparison with the second moment $\langle\phi^2\rangle$, $\sigma = \sqrt{\rho_s \epsilon} L_y / 2$. Expectation values in even/odd sectors $\mathcal{H}_{\epsilon,\phi}$ are given by the sums over even/odd ϕ with weight $P(\phi)$. As a result, differences between expectation values in $\mathcal{H}_{\epsilon,\phi}$ are exponentially small in L_y ; for example, the average squares of electric fluxes differ by $\langle\phi^2\rangle_e - \langle\phi^2\rangle_o \approx \exp(-\pi^2 \sqrt{\rho_s \epsilon} L_y / 4)$. An analysis of the disorder-induced level-splitting $\Delta_d \propto \exp(-L_x / \xi_1)$ at $t = v$ (in the triangular lattice) allows us to extract the correlation length $\xi_1 = \sqrt{3}/\ln(2 + \sqrt{3})$ along the cylinder axis.

Hence, the presence of diagonal dimers provides us with only two topological sectors for a system defined on a cylinder (four sectors on a torus); the ground states are degenerate in the limit $L_{x,y} \rightarrow \infty$, finite-size effects as well as local disorder lead to an exponentially weak energy splitting. In the thermodynamic limit the collapse of infinitely many to two topological sectors involves a quantum phase transition in the parameter μ . For $\mu = 0$ the above model corresponds to a symmetric triangular lattice where we expect favourable conditions for the construction of protected qubits.

Received 18 July; accepted 17 December 2001.

1. Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
2. Monroe, C., Meekhof, D., King, B., Itano, W. & Wineland, D. Demonstration of a fundamental quantum logic gate. *Phys. Rev. Lett.* **75**, 4714–4717 (1995).
3. Turchette, Q., Hood, C., Lange, W., Mabuchi, H. & Kimble, H. J. Measurement of conditional phase shifts for quantum logics. *Phys. Rev. Lett.* **75**, 4710–4713 (1995).
4. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
5. Shnirman, A., Schön, G. & Hermon, Z. Quantum manipulations of small Josephson junctions. *Phys. Rev. Lett.* **79**, 2371–2374 (1997).
6. Averin, D. V. Adiabatic quantum computation with Cooper pairs. *Solid State Commun.* **105**, 659–664 (1998).
7. Mooij, J. E. et al. Josephson persistent-current qubit. *Science* **285**, 1036–1039 (1999).
8. Ioffe, L., Geshkenbein, V. B., Feigel'man, M. V., Fauchère, A. L. & Blatter, G. Environmentally decoupled s-wave–d-wave–s-wave Josephson junctions for quantum computing. *Nature* **398**, 678–681 (1999).
9. Nakamura, Y., Pashkin, Yu. A. & Tsai, J. S. Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature* **398**, 786–788 (1999).
10. Friedman, J. R., Patel, V., Chen, W., Tolpygo, S. K. & Lukens, J. E. Quantum superposition of distinct macroscopic states. *Nature* **406**, 43–46 (2000).
11. van der Wal, C. H. et al. Quantum superposition of macroscopic persistent-current states. *Science* **290**, 773–777 (2000).
12. Kitaev, A. Yu. Fault-tolerant quantum computation by anyons. Preprint quant-ph/9707021 at xxx.lanl.gov (1997).
13. Preskill, J. In *Introduction to Quantum Computation and Information* (eds Lo, H.-K., Popescu, S. & Spiller, T.) 213–269 (World Scientific, Singapore, 1998).
14. Kivelson, S. A., Rokhsar, D. S. & Sethna, J. P. Topology of the resonating valence-bond state: solitons and high- T_c superconductivity. *Phys. Rev. B* **35**, 8865–8868 (1987).
15. Rokhsar, D. S. & Kivelson, S. A. Superconductivity and the quantum hard-core dimer gas. *Phys. Rev. Lett.* **61**, 2376–2379 (1988).
16. Wen, X. G. Mean-field theory of spin-liquid states with finite energy gap and topological orders. *Phys. Rev. B* **44**, 2664–2672 (1991).
17. Ioffe, L. B. & Larkin, A. I. Superconductivity in the liquid-dimer valence-bond state. *Phys. Rev. B* **40**, 6941–6947 (1989).
18. Anderson, P. W. The resonating valence bond state in La_2CuO_4 and superconductivity. *Science* **235**, 1196–1198 (1987).
19. Moessner, R. & Sondhi, S. L. Resonating valence bond phase in the triangular lattice quantum dimer model. *Phys. Rev. Lett.* **86**, 1881–1884 (2001).
20. Misguich, G., Lhuillier, C., Bernu, B. & Waldtmann, C. Spin-liquid phase of the multiple-spin exchange Hamiltonian on the triangular lattice. *Phys. Rev. B* **60**, 1064–1074 (1999).
21. Sachdev, S. Kagome-acute- and triangular-lattice Heisenberg antiferromagnets. *Phys. Rev. B* **45**, 12377–12396 (1992).
22. Grabert, H. & Devoret, M. H. *Single Charge Tunneling: Coulomb Blockade Phenomena in Nanos-structures* (Plenum, New York, 1992).
23. Fradkin, E. *Field Theories of Condensed Matter Systems* (Addison-Wesley, Redwood City, 1991).
24. Moessner, R., Sondhi, S. L. & Fradkin, E. Short-ranged RVB physics, quantum dimer models and Ising gauge theories. Preprint cond-mat/0103396 at xxx.lanl.gov (2001).
25. Fradkin, E. & Shenker, S. H. Phase diagrams of lattice gauge theories with Higgs fields. *Phys. Rev. D* **19**, 3682 (1979).

Acknowledgements

We thank V. Geshkenbein for discussions. We acknowledge financial support through the SCOPES programme (Swiss Federal Department of Foreign Affairs and Swiss National Foundation), the Dutch Organization for Fundamental Research (NWO), the Russian Foundation for Basic Research, the programme 'Quantum Macrophysics' of the Russian Academy of Science, and the Russian Ministry of Science. Computations were carried out on the Beowulfcluster Asgard at ETHZ.

Correspondence and requests for materials should be addressed to G.B. (e-mail: blatterj@itp.phys.ethz.ch).

Real-space imaging of an orbital Kondo resonance on the Cr(001) surface

O. Yu. Kolesnychenko, R. de Kort, M. I. Katsnelson, A. I. Lichtenstein & H. van Kempen

Research Institute for Materials, University of Nijmegen, Toernooiveld 1, NL-6525 ED Nijmegen, The Netherlands

The Kondo effect¹ is usually connected with the interaction between a localized spin moment and itinerant electrons. This interaction leads to the formation of a narrow resonance at the Fermi level, which is called the Abrikosov–Suhl or Kondo resonance². Scanning tunnelling microscopy is an ideal technique for real-space investigations of complicated electronic structures^{3,4} and many-body phenomena, such as the formation of the Kondo resonance^{5–8} or *d*-wave pairing in high-*T_c* superconductors⁹. Theory has predicted that similar, Kondo-like many-electron resonances are possible for scattering centres with orbital instead of spin degrees of freedom—the quadruple momenta in uranium-based compounds or two-level systems in metallic glasses are examples of such ‘pseudo-Kondo’ scattering centres¹⁰. Here we present evidence for the orbital Kondo resonance on a transition-metal surface. Investigations of an atomically clean Cr(001) surface at low temperature using scanning tunnelling microscopy reveal a very narrow resonance at 26 meV above the Fermi level, and enable us to visualize the orbital character of the corresponding state. The experimental data, together with many-body calculations, demonstrate that the observed resonance is an orbital Kondo resonance formed by two degenerate d_{xz} , d_{yz} surface states.

The experiments were performed on the Cr(001) surface at 4.2 K with a low-temperature scanning tunnelling microscope (STM) equipped with a sample cleaver¹¹. The Cr(001) surface was produced by *in situ* cleavage of a single crystal of 99.99% Cr, which had been prepared in the form of a bar 0.5 mm × 0.5 mm × 6 mm cut along the {001} crystallographic direction by spark erosion. As we found out, the cleavage of chromium single crystals produces atomically flat and clean {001} surfaces.

The electronic structure with atomic spatial resolution was investigated by scanning tunnelling spectroscopy (STS). Figure 1 shows a typical dI/dV spectrum measured on an atomically clean Cr(001) surface in the middle of a large monoatomic terrace. A narrow resonance at 26 meV above the Fermi level can be seen. The resonance was very well reproduced (in terms of position and width) on different surfaces produced by fracture of four samples, and when using different PtIr tips. (The tip is scanned across the surface in this technique.) Therefore, we can conclude that the observed resonance is an intrinsic property of the Cr(001) surface, and is not related to stress—which might be introduced into the crystal during the fracture of the sample—or to tip properties.

The orbital symmetry of this resonance can be revealed by monitoring the suppression of this surface state by a single impurity, in analogy, for example, to the visualization of a superconducting gap near a zinc atom⁹. An impurity locally destroys a coherent state, and allows the visualization of the character and spatial symmetry of this state with the atomic-scale STM probe^{9,12}. Figure 2a shows an STM image of a Cr(001) surface with various types of impurities in the first layer (and in the deeper, buried layers), obtained at a bias voltage V_b far from the resonance. The charge density oscillations¹³ set up around the impurities in order to screen the local potential can be seen. This oscillation pattern persists for the

whole studied energy range. In addition, in the narrow energy region ($10 \text{ meV} < eV_b < 50 \text{ meV}$) corresponding to the resonance, all STM images show cross-like depressions centred around the impurities (Fig. 2b). The size of this feature is about 4.5 nm in the {100} directions. We found two main types of impurities on the cleaved surface: one is seen as a protrusion (for example, 2 in Fig. 2b), corresponding probably to S, and another is seen as a depression (for example, 1 in Fig. 2b), corresponding probably to N, O or C, as these elements are the most common impurities for a Cr single crystal¹⁴. Although the exact origin of the point defects on the Cr(001) surface is unknown, we note that the cross-like feature appears on both types of impurities and is related to the intrinsic electronic properties of the Cr(001) surface. Therefore this cross-like feature does not depend on the impurity species which just produces a symmetry-breaking perturbation. From the line profile (Fig. 2c) between A and B in Fig. 2b, we note that even inside the depression atomic corrugation can be seen, as indicated by arrows. This depression is the result of a decrease of the density of states at the energy corresponding to the resonance with cross-like spatial distribution, as we confirmed by dI/dV spectroscopic measurements. Taking into account the symmetry of the observed cross-like feature, we can conclude that the surface state corresponding to the observed resonance has d_{xz} , d_{yz} orbital symmetry. The other surface states that can exist on the Cr(001) surface can be excluded: d_{z^2} , for example, would result in a circle-like feature due to its in-plane symmetry, whereas $d_{x^2-y^2}$ and d_{xy} orbitals cannot form localized surface states because of their strong overlapping.

The resonance in the density of states on the Cr(001) surface was first observed at room temperature³, and was ascribed to a one-electron d_{z^2} surface state. However, theoretical calculations^{3,15,16} showed that the d_{z^2} surface state is located around 550 meV above the Fermi level, and only drastic reduction of the surface magnetic moments could bring this state close to the Fermi level³. Even in such a case, the d_{z^2} state, being of the wrong orbital symmetry, cannot be responsible for the cross-like feature. The one-electron surface states required by d_{xz} , d_{yz} symmetry are located far away from

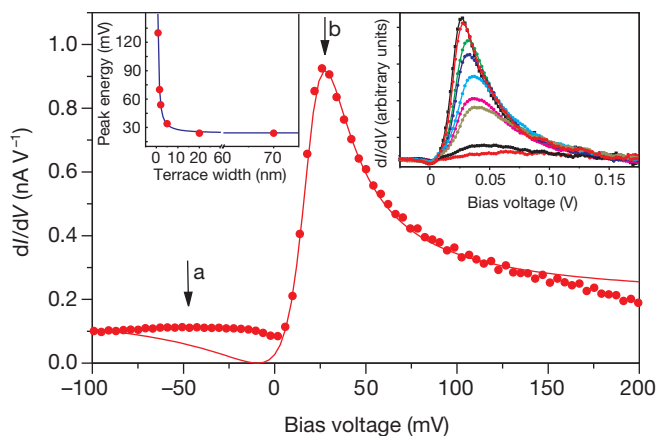


Figure 1 Tunnelling spectroscopy on Cr(001). The main figure shows the averaged tunnelling conductance versus sample bias voltage, measured in the middle of a Cr(001) terrace of 70 nm width. The tunnel current I_t versus V_b curve was obtained using 200 points in the energy window of $\pm 200 \text{ mV}$. The red solid curve shows a fit by the Fano shape⁵ with $T_K = 15 \text{ meV}$, $q = 1.9$ and $\alpha = 20 \text{ meV}$. The left inset shows the dependence of the resonance energy on the terrace width measured in the middle of a terrace. The drawn curve shows the fit by a $A + B/L^{1.3}$ function. The right inset presents position-dependent dI/dV spectra (the nine curves from bottom to top were taken in steps of 0.25 nm from the step edge to the terrace centre). Two arrows marked a and b indicate bias voltages corresponding to STM images presented in Fig. 2a and b, respectively.

the Fermi level¹⁶. The present experimental results show that the observed narrow state on the Cr(001) surface is a many-electron orbital Kondo resonance.

Tight-binding mean-field calculations for the ferromagnetic Cr(001) surface¹⁶ demonstrated the existence of two degenerate spin-split d_{xz} , d_{yz} surface states, located about 1 eV below and above the Fermi level. The spatial symmetry of these states on the (001) surface protects the degeneracy, and the interaction of such states with the conduction electrons can lead to the formation of a many-body Kondo resonance near the Fermi level. The d_{xz} state, being hybridized with the conduction electron bands, changes into the d_{yz} state via virtual many-body excitations and vice versa (see Fig. 3, right inset). In this process, a many-body 'orbital singlet' can be formed with the d_{xz} , d_{yz} states mixed symmetrically in the surface plane, similar to the spin-flip mechanism of the local magnetic moment screening in the conventional Kondo effect¹⁰. These orbital flips lead to the formation of a many-electron resonance at the Fermi energy—the orbital Kondo resonance. Because we

have a whole two-dimensional lattice of resonant scattering centres (every surface atom), a narrow quasiparticle band is formed near the Fermi energy, similar to the formation of the heavy-fermion state in the spin Kondo lattices². It is important to stress that we consider the orbital Kondo resonance due to a double degeneracy of the d_{xz} , d_{yz} states and not the spin Kondo resonance, as the spin degeneracy is broken by the ferromagnetism of the Cr(100) surface¹⁷.

There are several reasons to prefer the many-electron interpretation over a one-electron model in the context of our experimentally observed results. First, the resonance is very narrow, and one-electron d_{xz} , d_{yz} surface states could hardly give such a narrow peak in the density of states. The typical width of any one-electron surface state is of the order of several hundreds of meV (refs 3, 15, 16). In contrast, the width of the Kondo resonance is determined by the Kondo temperature (T_K ; refs 2 and 5) and can be very small¹⁰. A fit by the Fano shape⁵ to the spectrum in Fig. 1 gives $T_K = 15$ meV. In agreement with expectations^{2,10} that the Kondo resonance will be shifted and logarithmically broadened at $T > T_K$, room temperature measurements³ show a broad peak near the Fermi level. Second, in any one-electron model (for example, for d_{z^2}), a small width of the peak means a strong localization in real space. This is in contradiction to the experimentally observed large spatial influence of the impurities (the size of the cross-like feature is of the order of ten interatomic distances). In contrast, for Kondo lattices the resonance of many-electron origin is formed by the states from a very narrow region in k -space, and therefore the effect of an impurity on this resonance exhibits a large extension in real space².

There is an essential difference between our STS data, concerning the orbital Kondo lattice, and the previous experimental results on the spectroscopy of Kondo resonances for isolated magnetic impurities^{5–7}. In the orbital Kondo lattice, the d_{xz} , d_{yz} surface states contribute directly to the current, which results in the resonance

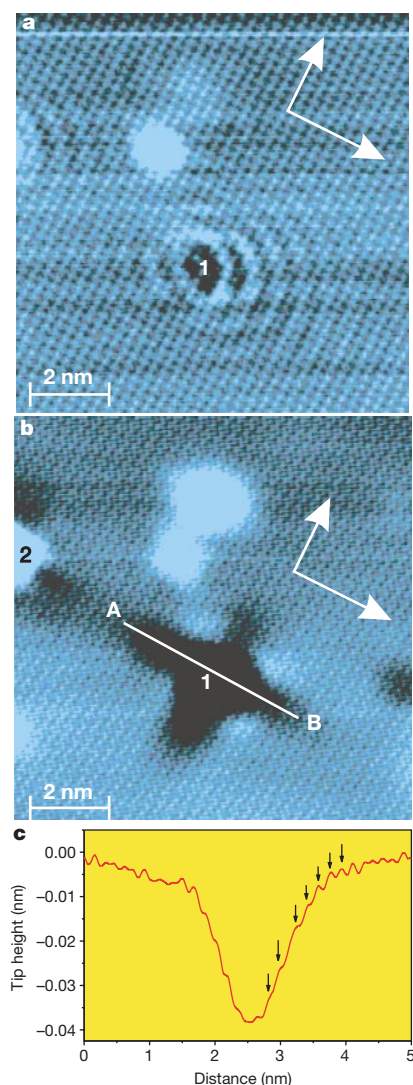


Figure 2 Visualization of the orbital Kondo resonance on Cr(001). **a**, 10 nm × 10 nm STM image obtained at $I_t = 0.5$ nA and $V_b = -50$ mV, showing charge density oscillations set up around impurities. **b**, 10 nm × 10 nm STM image obtained at $I_t = 0.5$ nA and $V_b = 25$ mV, showing a cross-like depression centred around impurities. **c**, The line profile between A and B in **b**. The arrows indicate some positions of atom protrusions. See text for details of features 1 and 2.

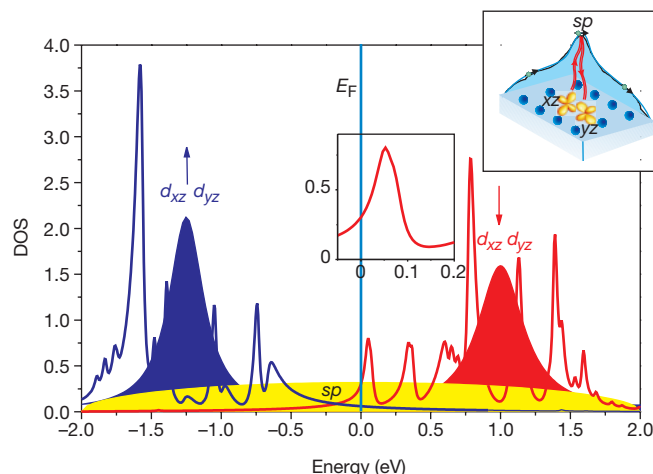


Figure 3 Formation of the orbital Kondo resonance. Main panel, spin-polarized density of states (DOS) for a two-degenerate periodic Anderson model. Zero energy on the horizontal axis corresponds to the Fermi energy, E_F . The bare DOS for the ferromagnetic Cr(001) surface is simulated by the semicircular model for a conducting sp band of 4 eV width (yellow) and one-electron d_{xz} , d_{yz} spin-up (blue) and spin-down (red) surface states¹⁶. The effective hybridization of 0.4 for spin-up and 0.5 eV for spin-down interactions together with Coulomb parameters $U = 1.2$ eV and $J = 0.4$ eV were used. The corresponding DOS for correlated electrons is shown by blue and red curves. Left inset, the orbital Kondo resonance on a larger scale. Right inset, a pictorial view of the orbital Kondo resonance formation on the Cr(001) surface. A virtual transition between d_{xz} and d_{yz} surface states via hybridization with the conduction sp band leads to the formation of an orbital singlet state.

in the dI/dV spectra. A similar effect is well known in heavy-fermion physics, where heavy f -electrons contribute to superconductivity, the de Haas–van Alfen effect, and so on (ref. 2). We note that the shape and the width of the Cr(001) resonance is similar to the Kondo resonance observed on Co clusters⁸.

Quantum state degeneracy of a scattering centre is crucial for the existence of the Kondo effect. In the case of the orbital Kondo resonance, the surface-symmetry-protected degeneracy can be destroyed by a step edge. The potential created by a step edge destroys the x, y symmetry in the surface plane, breaking the degeneracy of (that is, splitting) the d_{xz} and d_{yz} surface states. Our experimental results (right inset in Fig. 1) do show a suppression of the Kondo resonance (shifting and broadening) on the approach to a step edge. In the framework of the orbital Kondo model, the influence of a monoatomic terrace should be similar to the influence of a weak magnetic field on the conventional spin Kondo resonance². In the extensively studied case of single magnetic impurities, the magnetic field splits the Kondo resonance². But for the general non-symmetric Kondo lattice problem, the shifting and broadening of the many-body resonance is a more general way of destroying the Kondo coherence. The dependence of the d_{xz} – d_{yz} energy splitting—or an effective ‘pseudomagnetic field’ $h_{\text{eff}}^{\text{orb}}$ —on the distance from the step edge L can be estimated in first-order perturbation theory to within the change of the total electron density due to the potential of the step edge. This results from the Friedel oscillations on the surface, which lead to $h_{\text{eff}}^{\text{orb}} \propto L^{-3/2}$ for a general shape of the Fermi surface¹³, and to $h_{\text{eff}}^{\text{orb}} \propto L^{-1}$ for the case of a pronounced ‘nesting’ on the Fermi surface. Our experimental data show that the shift of the resonance is fitted by a $L^{-1.3 \pm 0.2}$ law, in agreement with the theoretical explanation.

In order to verify quantitatively the formation of the orbital Kondo resonance on the Cr(001) surface, we have performed calculations for the periodic degenerate Anderson model^{2,18}. We use the dynamical mean-field theory (for a review, see ref. 18), which was used for the investigation of the electronic structure for a one-band periodic Anderson model¹⁹. The results of the calculations are shown in the main panel of Fig. 3. The orbital Kondo resonance (Fig. 3, left inset), in the vicinity of the Fermi level, forms under very general conditions. A reasonably small intrinsic bandwidth for d_{xz} , d_{yz} states cannot destroy it, as the Kondo resonance is also known for the itinerant-electron Hubbard model¹⁸. The characteristics of spectra for larger energies ($|E - E_F| > 0.3$ eV) are more sensitive to the details of the real conduction band structure.

In this work, we have shown that the orbital degeneracy of the d_{xz} and d_{yz} surface states can lead to the formation of an orbital Kondo resonance on Cr(001). We expect that the surfaces of other transition metals (for example, manganese) will show similar features, provided that an orbital-degenerate surface state exists. □

Received 12 September; accepted 3 December 2001.

- Kondo, J. Resistance minimum in dilute magnetic alloys. *Prog. Theor. Phys.* **32**, 37–49 (1964).
- Hewson, A. C. *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, Cambridge, 1993).
- Strosio, J. A., Pierce, D. T., Davies, A., Celotta, R. J. & Weinert, M. Tunneling spectroscopy of bcc (001) surface states. *Phys. Rev. Lett.* **75**, 2960–2963 (1996).
- Heinze, S. *et al.* Real space imaging of two-dimensional antiferromagnetism on the atomic scale. *Science* **288**, 1805–1808 (2000).
- Madhavan, V., Chen, W., Jamneala, T., Crommie, M. F. & Wingreen, N. S. Tunneling into a single magnetic atom. *Science* **280**, 567–569 (1998).
- Manoharan, H. C., Lutz, C. P. & Eigler, D. M. Quantum mirages formed by coherent projection of electronic structure. *Nature* **403**, 512–515 (2000).
- Li, J., Schneider, W.-D., Berndt, R. & Delley, D. Kondo scattering observed at a single magnetic impurity. *Phys. Rev. Lett.* **80**, 2893–2896 (1998).
- Odom, T. W., Huang, J.-L., Cheung, C. L. & Lieber, C. M. Magnetic clusters on single-walled carbon nanotubes: the Kondo effect in a one-dimensional host. *Science* **290**, 1549–1552 (2000).
- Pan, S. H. *et al.* Imaging the effects of individual zinc impurity atoms on superconductivity in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Nature* **403**, 746–750 (2000).
- Cox, D. L. & Zawadowski, A. Exotic Kondo effects in metals: magnetic ions in a crystalline electric field and tunneling centers. *Adv. Phys.* **47**, 599–942 (1998).

- van der Wielen, M. C. M. M., van Roij, A. J. A. & van Kempen, H. Direct observation of Friedel oscillations around Incorporated Si_{Ga} dopants in GaAs by low-temperature scanning tunneling microscopy. *Phys. Rev. Lett.* **76**, 1075–1078 (1996).
- Balatsky, A. V. Superconductivity: from obscurity to impurity. *Nature* **403**, 717–718 (2000).
- Crommie, M. F., Lutz, C. P. & Eigler, D. M. Imaging standing waves in a two-dimensional electron gas. *Nature* **363**, 524–527 (1993).
- Schmid, M., Pinczolits, M., Hebenstreit, W. & Varga, P. Segregation of impurities on Cr(001) studied by AES and STM. *Surf. Sci.* **377**, 1023–1027 (1997).
- Fu, C. L. & Freeman, A. J. Surface ferromagnetism of Cr(001). *Phys. Rev. B* **33**, 1755–1761 (1986).
- Klebanoff, L. E., Vitoria, R. H., Falicov, L. M. & Shirley, D. A. Experimental and theoretical investigations of Cr(001) surface electronic structure. *Phys. Rev. B* **32**, 1997–2005 (1985).
- Kleiber, M., Bode, M., Ravlic, R. & Wiesendanger, R. Topology-induced spin frustration at the Cr(001) surface studied by spin-polarized scanning tunneling spectroscopy. *Phys. Rev. Lett.* **85**, 4606–4609 (2000).
- Georges, A., Kotliar, G., Krauth, W. & Rozenberg, M. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Mod. Phys.* **68**, 13–125 (1996).
- Jarrell, M. Symmetrical periodic Anderson model in infinite dimensions. *Phys. Rev. B* **51**, 7429–7440 (1995).

Acknowledgements

We thank J. Hermesen and J. Gerritsen for technical assistance, and A. Keen for a reading of the manuscript. This work was supported by FOM and NWO.

Correspondence and requests for materials should be addressed to H.v.K. (e-mail: hvk@sci.kun.nl).

Stepwise radial complexation of imine groups in phenylazomethine dendrimers

Kimihisa Yamamoto, Masayoshi Higuchi, Satoshi Shiki, Masanori Tsuruta & Hiroshi Chiba

Kanagawa Academy of Science & Technology (KAST), and Department of Chemistry, Faculty of Science & Technology, Keio University, Yokohama 223-8522, Japan

Dendrimers^{1–6} are highly branched organic macromolecules with successive layers or ‘generations’ of branch units surrounding a central core. Organic–inorganic hybrid versions have also been produced, by trapping metal ions or metal clusters within the voids of the dendrimers^{7–13}. The unusual, tree-like topology endows these nanometre-sized macromolecules with a gradient in branch density from the interior to the exterior, which can give rise to an energy gradient that directs the transfer of charge and energy from the dendrimer periphery to its core^{4–6}. Here we show that tin ions, Sn^{2+} , complex to the imine groups of a spherical polyphenylazomethine dendrimer in a stepwise fashion. This behaviour reflects a gradient in the electron density associated with the imine groups, with complexation in a more peripheral generation proceeding only after complexation in generations closer to the core has been completed. By attaching an electron-withdrawing group to the dendrimer core, we are able to change the complexation pattern, so that the core imines are complexed last. By further extending this strategy, it should be possible to control the number and location of metal ions incorporated into dendrimer structures, which might find uses as tailored catalysts or building blocks for advanced materials.

The π -conjugated dendritic polyphenylazomethines (DPAs G1–4, designated as GX, where X is the generation number) were synthesized using a convergent strategy^{14,15}. DPA G4 (Fig. 1a) has a sphere-like structure (diameter 2.3 nm) and exhibits conformational rigidity, as evidenced by the lack of molecular deformation after deposition on plates¹⁵. The imine sites present in the DPAs

coordinate strongly to various metal ions, while SnCl_2 is a Lewis acid forming 1:1 complexes with imine compounds, and complexation of the 30 imine sites in DPA G4 (Fig. 1a) should therefore yield a nanosized spherical complex. Addition of SnCl_2 to a dichloromethane/acetonitrile solution of DPA G4 resulted in a colour change from yellow to orange, attributed to complexation. We found that complexation was complete in about 10 minutes, as determined from the spectral changes.

Using UV-visible spectroscopy to monitor the titration until an equimolar amount of SnCl_2 , we observed four changes in the position of the isosbestic point (Fig. 2a), indicating that the complexation proceeds not randomly, but stepwise. An isosbestic point appears when a compound is quantitatively transformed into another by complexation¹⁶, so the four different isosbestic points we observe suggest that four different complexes are successively formed upon SnCl_2 addition. The spectra of DPA G4 gradually changed, with an isosbestic point at 375 nm up to the addition of 2 equivalents of SnCl_2 (Fig. 2b). The isosbestic point then shifted upon further addition of SnCl_2 , and appeared at 364 nm between 3 and 6 equivalents (Fig. 2c). While adding between 7 and 14 equivalents of SnCl_2 , an isosbestic point appeared at 360 nm, moving to 355 nm when adding between 15 and 30 equivalents (Fig. 2d and e). Overall, the isosbestic point shifted about 20 nm from 375 to 355 nm, and the number of added equivalents of SnCl_2 required to induce a shift was in agreement with the number of imine sites present in the different generations of DPA G4.

From a kinetic standpoint, complexation of the terminal imines of the dendrimer is expected to occur first. But the titration results suggest that on the timescale of our observations the process is thermodynamically controlled and proceeding in a stepwise fashion from the core imines to the terminal imines of DPA G4 (Fig. 3). Similar stepwise complexation was also observed with DPA G2 and G3 (see Figs 2 and 3 in Supplementary Information). For DPA G2, two isosbestic points appeared—at 344 and 355 nm—when adding 0–2 and 3–6 equivalents of SnCl_2 , respectively. In the case of DPA G3, three isosbestic points centred at 367, 360 and 355 nm appeared

when adding between 0–2, 3–6 and 7–14 equivalents of SnCl_2 , respectively. Again, the equivalents of SnCl_2 added before a shift in isosbestic point is observed agree with the number of imine sites present in the different generations of the two dendrimers. These results further supported the idea that metal ions are incorporated in a stepwise fashion, filling first the generations close to the dendrimer core and then progressively more peripheral generations. (See Fig. 7 in Supplementary Information for information on the complexation behaviour involving other metal salts.)

The spectral changes accompanying the addition of SnCl_2 to the G4 dendron (Fig. 1b) and the linear oligomers **1** and **2** (Fig. 1c and d)^{17,18} resulted in each case in the appearance of only one isosbestic point (Fig. 4, and also Figs 4 and 5 in Supplementary Information). This behaviour is expected if the complexation between the metal ions and all available imine sites is random. These observations therefore suggest that the stepwise complexation behaviour seen with the dendrimers does not reflect changes in the basicity of imine sites induced by complexation at neighbouring imine sites. If that were the case, we would expect to also see stepwise complexation when titrating the G4 dendron and oligomers **1** and **2**.

The stepwise complexation behaviour of the dendrimers must therefore result from differences in the basicity of the imines and/or differences in accessibility to the amine sites in the different generations due to steric hindrance. The experimental observations suggest that the basicity of the core imines is higher than that of the more peripheral imines. This conclusion is further supported by determinations of complexation constants K using UV-visible

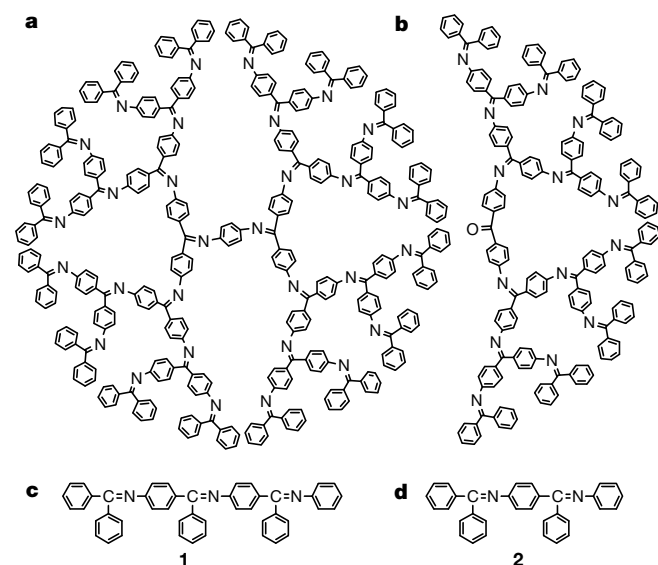


Figure 1 Chemical structures of the dendrimers and oligomers investigated by complexation with SnCl_2 . **a**, Dendritic polyphenylazomethine G4 (DPA G4, designated as GX, where X is the generation number). **b**, DPA dendron G4. **c**, **d**, The linear oligomers **1** and **2**. For the synthesis of DPA G4 and the dendron, see refs 14 and 15. DPA G4 has a sphere-like structure (diameter 2.3 nm). For the synthesis of the linear oligomers **1** and **2**, see refs 17 and 18.

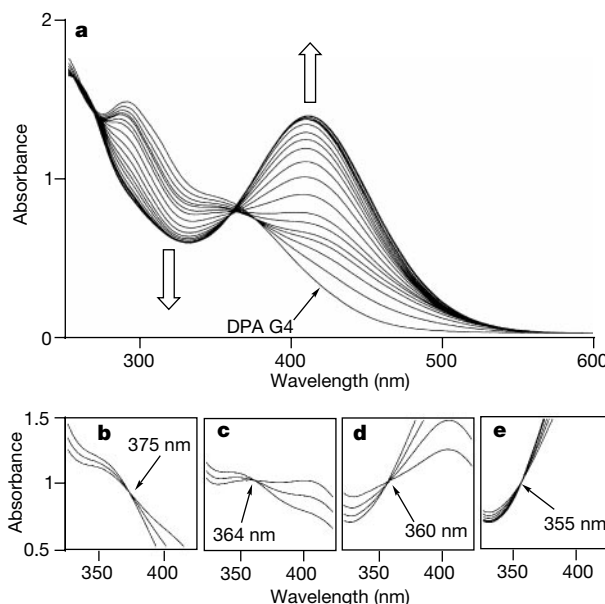


Figure 2 UV-visible spectra of DPA G4 complexed with SnCl_2 . The DPA G4 was dissolved in 1:1 dichloromethane/acetonitrile at a concentration of 5.0×10^{-6} M. **a**, The spectral change during the addition of 0–30 equivalents of SnCl_2 . Four changes in the isosbestic point were observed during the titration. **b**, The spectral change during the addition of 0–2 equivalents of SnCl_2 (isosbestic point at 375 nm). **c**, The spectral change during the addition of 3–6 equivalents of SnCl_2 (isosbestic point at 364 nm). **d**, The spectral change during the addition of 7–14 equivalents of SnCl_2 (isosbestic point at 360 nm). **e**, The spectral change during the addition of 15–30 equivalents of SnCl_2 (isosbestic point at 355 nm). The four changes in the isosbestic point show that four different complexes are successively formed during the SnCl_2 addition. The equivalents of SnCl_2 agree with the number of imine sites for each generation of the dendrimers. These results supported the idea that metal ions are assembled from the core side to the outside in the interior of the dendrimer as shown in Fig. 3.

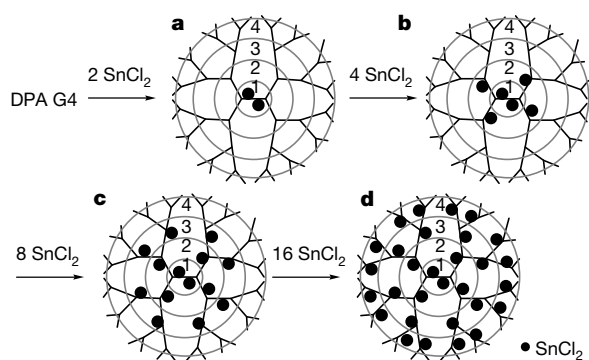


Figure 3 Schematic representation of stepwise complexation of DPA G4 with SnCl_2 . DPA G4 complexed with **a**, 2, **b**, 4, **c**, 8, and **d**, 16 equivalents of SnCl_2 . This complexation thermodynamically proceeds in a stepwise fashion from the core imines to the terminal ones of DPA G4 based on the electron gradients. Similar stepwise complexation was also confirmed in DPA G2 and G3 (see Figs 2 and 3 in Supplementary Information).

spectroscopy. We measured values of $1.4 \times 10^4 \text{ M}^{-1}$ for DPA G1, which serves as a model for the core generation (Fig. 1 in Supplementary Information), and a lower value of $0.96 \times 10^4 \text{ M}^{-1}$ for the linear oligomer **2**, which serves as a model for a terminal dendrimer branch. ^{13}C NMR measurements show that the peak attributed to imine carbon in the dendrimer core is shifted by more than 1 p.p.m. towards the high magnetic field, relative to the peak attributed to imine carbon at the periphery, further confirming that DPA core imines are more basic. Steric hindrance could also explain the stepwise complexation behaviour because steric crowding of the terminal branches increases with generation number¹⁹ and could render complexation of terminal imines more difficult.

The complexation of DPAs with SnCl_2 was confirmed by atomic absorption spectroscopy and infrared spectroscopy (see Supplementary Information). The stepwise complexation of DPAs was not only observed during the UV–visible spectral measurements, but also using three other approaches: (1) monitoring the change of the ^1H NMR spectrum of DPA G2 during the complexation with SnCl_2 (Fig. 8 in Supplementary Information), (2) direct imaging of DPA G4 molecules during different complexation stages by transmission

electron microscopy (Fig. 9 in Supplementary Information), and (3) direct confirmation of the stepwise complexation with SnCl_2 by shell-selective reduction of imines in the metallodendrimers (Figs 10, 11 and 12 in Supplementary Information).

We also changed the core phenyl ring of DPA G3 to tetrafluorophenyl, an electron-withdrawing group, by dehydration of the G3 dendron with tetrafluoro-1,4-phenylenediamine (see Supplementary Information for experimental procedures and characterization data). When adding SnCl_2 to the modified DPA G3 carrying a tetrafluorophenyl core (DPA-F G3), three isosbestic points appeared at 359, 366 and 369 nm upon adding 0–4, 5–12 and 13–14 equivalents of SnCl_2 , respectively (Fig. 6 in Supplementary Information). That is, the imines present in the second generation were complexed first, then those present in the third generation, and finally those present in the first (core) generation, whose basicity was lowered by the electron-withdrawing tetrafluorophenyl group. This result illustrates that it is possible to control the basicity gradient in dendrimers by attaching electron-withdrawing or electron-donating substituents to the dendrimer core. Similar effects should occur when attaching suitable groups at the dendrimer periphery. □

Received 22 June; accepted 6 December 2001.

- Tomalia, D. A., Dewald, J., Hall, M., Martin, S. & Smith, P. B. in *Preprints 1st SPSJ Int. Polym. Conf.* (ed. Uematsu, I.) 65 (Society of Polymer Science, Japan, Tokyo, 1984).
- Tomalia, D. A. et al. A new class of polymers: starburst-dendritic macromolecules. *Polym. J.* **17**, 117–132 (1985).
- Newkome, G. R., Yao, Z. Q., Baker, G. R. & Gupta, V. K. Cascade molecules: a new approach to micelles. *A [27]-arborol. J. Org. Chem.* **50**, 2003–2004 (1985).
- Devadoss, C., Bharathi, P. & Moore, J. S. Energy transfer in dendritic macromolecules: molecular size effects and the role of an energy gradient. *J. Am. Chem. Soc.* **118**, 9635–9644 (1996).
- Jiang, D. L. & Aida, T. Photoisomerization in dendrimers by harvesting of low-energy photons. *Nature* **388**, 454–456 (1997).
- Selby, T. D. & Blackstock, S. C. Preparation of a redox-gradient dendrimer. Polyamines designed for one-way electron transfer and charge capture. *J. Am. Chem. Soc.* **120**, 12155–12156 (1998).
- Balogh, L. & Tomalia, D. A. Poly(amidoamine) dendrimer-templated nanocomposites. 1. Synthesis of zerovalent copper nanoclusters. *J. Am. Chem. Soc.* **120**, 7355–7356 (1998).
- Zhao, M., Sun, L. & Crooks, R. M. Preparation of Cu nanoclusters within dendrimer templates. *J. Am. Chem. Soc.* **120**, 4877–4878 (1998).
- Crooks, R. M., Zhao, M., Sun, L., Chechik, V. & Yeung, L. K. Dendrimer-encapsulated metal nanoparticles: synthesis, characterization, and applications to catalysis. *Acc. Chem. Res.* **34**, 181–190 (2001).
- Tomimaga, M., Hosogi, J., Konishi, K. & Aida, T. Nucleobase dendrimer as a multidentate ligand for a rare-earth metal ion. *Chem. Commun.* 719–720 (2000).
- Diaz, D. J., Storrier, G. D., Bernhard, S., Takada, K. & Abruna, H. D. Ordered arrays generated via metal-initiated self-assembly of terpyridine containing dendrimers and bridging ligands. *Langmuir* **15**, 7351–7354 (1999).
- Petrucchi-Samija, M., Guillemette, V., Dasgupta, M. & Kakkar, A. K. A new divergent route to the synthesis of organophosphine and metallodendrimers via simple acid-base hydrolytic chemistry. *J. Am. Chem. Soc.* **121**, 1968–1969 (1999).
- Kawa, M. & Fréchet, J. M. J. Self-assembled lanthanide-cored dendrimer complexes: enhancement of the luminescence properties of lanthanide ions through site-isolation and antenna effects. *Chem. Mater.* **10**, 286–296 (1998).
- Higuchi, M., Shiki, S. & Yamamoto, K. Novel phenylazomethine dendrimers: synthesis and structural properties. *Org. Lett.* **2**, 3079–3082 (2000).
- Higuchi, M., Shiki, S., Ariga, K. & Yamamoto, K. First synthesis of phenylazomethine dendrimer ligands and structural studies. *J. Am. Chem. Soc.* **123**, 4414–4420 (2001).
- Schmidt, E., Zhang, H., Chang, C. K., Babcock, G. T. & Oertling, W. A. Room temperature binding of CO to cobaltous porphyrin p cation radical: spectroscopic characterization of mono and bis CO complexes with cobaltic porphyrin. *J. Am. Chem. Soc.* **118**, 2954–2961 (1996).
- Higuchi, M. & Yamamoto, K. Novel cyclic molecules: selective synthesis of cyclic phenylazomethines. *Org. Lett.* **1**, 1881–1883 (1999).
- Higuchi, M., Kimoto, K., Shiki, S. & Yamamoto, K. Selective synthesis of novel cyclic phenylazomethine trimers. *J. Org. Chem.* **65**, 5680–5684 (2000).
- Hecht, S. & Fréchet, J. M. J. An alternative synthesis approach toward dendritic macromolecules: novel benzene-core dendrimers via alkyne cyclotrimerization. *J. Am. Chem. Soc.* **121**, 4084–4085 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

This work was partially supported by a Grant-in-Aid for priority area and for scientific research from the Ministry of Education, Culture, Sports, Science & Technology Foundation, a Grant-in-Aid for evaluative technology from the Science and Technology Agency, and a Kanagawa Academy science and technology research grant.

Correspondence and requests for materials should be addressed to K.Y. (e-mail: yamamoto@chem.keio.ac.jp).

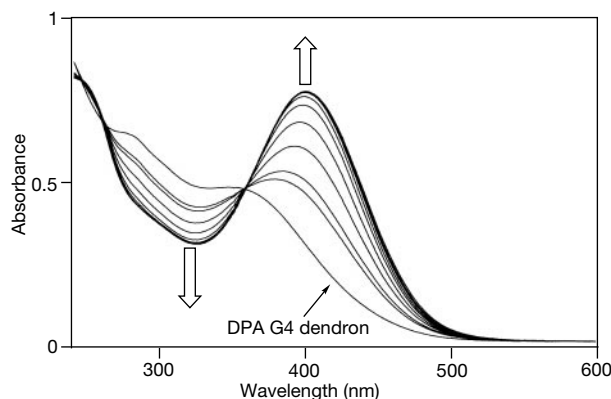


Figure 4 UV–visible spectra of DPA G4 dendron during complexation with SnCl_2 . Shown is the complexation of the dendron while adding 0–14 equivalents of SnCl_2 . DPA G4 dendron was dissolved in 1:1 dichloromethane:acetonitrile at a concentration of $5 \times 10^{-6} \text{ M}$. The spectra show only one isosbestic point throughout the addition of SnCl_2 . The linear oligomers **1** and **2** also showed only one isosbestic point with the addition of SnCl_2 (Figs 4 and 5 in Supplementary Information). These results show that SnCl_2 molecules are randomly complexed to the imine sites of these compounds.

Quantifying the risk of extreme seasonal precipitation events in a changing climate

T. N. Palmer* & J. Räisänen†

* European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berks RG2 9AX, UK

† Rosby Centre, SMHI, S-60176 Norrköping, Sweden

Increasing concentrations of atmospheric carbon dioxide will almost certainly lead to changes in global mean climate¹. But because—by definition—extreme events are rare, it is significantly more difficult to quantify the risk of extremes. Ensemble-based probabilistic predictions², as used in short- and medium-term forecasts of weather and climate, are more useful than deterministic forecasts using a ‘best guess’ scenario to address this sort of problem^{3,4}. Here we present a probabilistic analysis of 19 global climate model simulations with a generic binary decision model. We estimate that the probability of total boreal winter precipitation exceeding two standard deviations above normal will increase by a factor of five over parts of the UK over the next 100 years. We find similar increases in probability for the Asian monsoon region in boreal summer, with implications for flooding in Bangladesh. Further practical applications of our techniques would be helped by the use of larger ensembles (for a more complete sampling of model uncertainty) and a wider range of scenarios at a resolution adequate to analyse average-size river basins.

The autumn and winter of 2000/2001 were the wettest on record over England and Wales, with widespread flooding⁵. Throughout the period, there was a general sense of concern that such heavy and prolonged precipitation was attributable in some measure to anthropogenic global warming. In attempting to address such societal concerns, we pose the question: how does anthropogenic forcing influence the probability of occurrence of unusually large seasonal precipitation amounts?

Specifically, we consider the dichotomous event E_n , defined to occur if the total seasonal precipitation at a specific location exceeds n standard deviations (σ) above the mean (μ). (Here μ and σ are location-dependent and seasonally dependent climate statistics, associated with twentieth-century levels of atmospheric CO_2 .) We restrict attention to $n = 2$ and $n = 3$, and use the conventional meteorological delineation (December–February, March–May, June–August and September–November) of seasons.

The methodology of ref. 6, based on the risk of any dichotomous climate event E , is used to analyse the changing probability of the extreme events E_2 and E_3 , as determined by an ensemble of climate projections. Here we use 80-year integrations from the CMIP2 (second coupled model intercomparison project) multi-model ensemble of 19 global coupled ocean–atmosphere climate models⁷ as discussed in the recent IPCC assessment¹. The benefit of the multi-model ensemble (over a single-model ensemble) accrues from its sampling some of the inevitable uncertainties in the computational representation of the equations of climate⁸. The first (control) ensemble was run with a constant twentieth-century CO_2 concentration (about 330 p.p.m.v.), and the second (greenhouse) ensemble with a transient compound increase in CO_2 of 1% per year. This increase in CO_2 is somewhat faster than is anticipated for the twenty-first century, but its use can be justified from the neglect of other anthropogenic greenhouse gases in CMIP2. More specifically, the global-mean radiative forcing due to the 70-year doubling of CO_2 in the CMIP2 experiments (3.7 W m^{-2}) is in the mid-range of the IPCC projections of the change in radiative forcing

from 1990 to 2060 ($2.6\text{--}5.0 \text{ W m}^{-2}$ for different forcing scenarios). Although it would be desirable to repeat this analysis based on other trace-gas scenarios, sufficiently large multi-model ensembles with alternative scenarios do not at present exist.

The control CMIP2 ensemble is used to define a ‘baseline’ twentieth-century probability of E_n , whilst the probability of E_n at the time of CO_2 doubling is estimated from years 61–80 of the CMIP2 greenhouse ensemble. Figure 1a shows the probability of

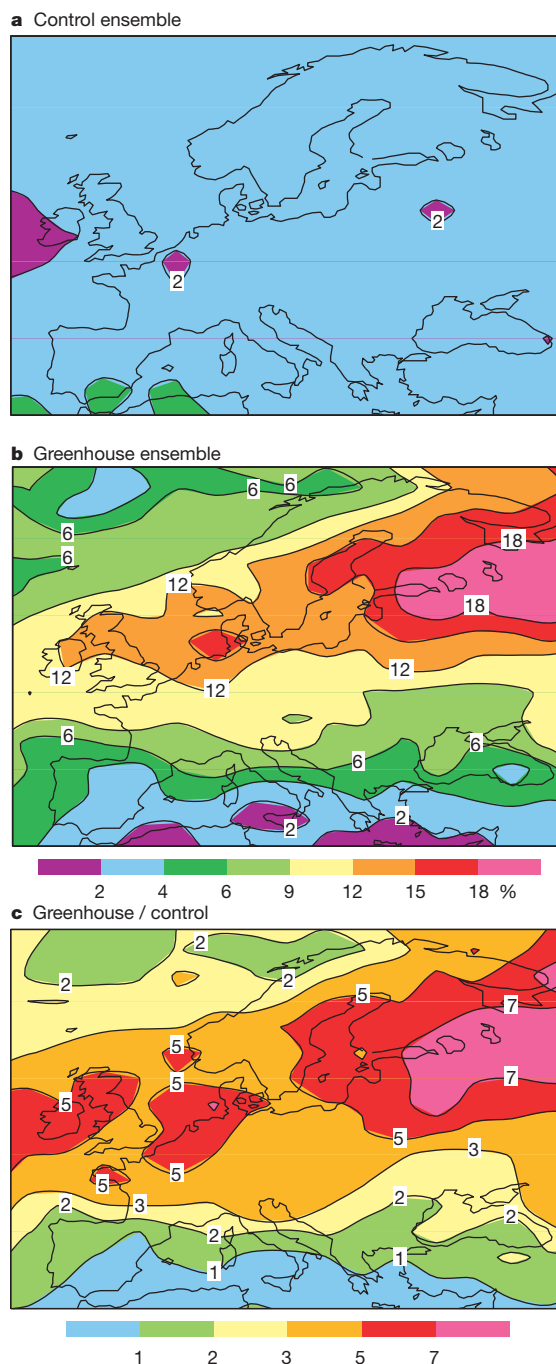


Figure 1 The changing probability of extreme seasonal precipitation for Europe in boreal winter. **a**, The probability (in %) of a ‘very wet’ winter defined from the control CMIP2 ensemble with twentieth-century levels of CO_2 and based on the event E_2 : total boreal winter precipitation greater than the mean plus two standard deviations. **b**, The probability of E_2 but using data from the greenhouse ensemble with transient increase in CO_2 , and calculated around the time of CO_2 doubling (years 61–80 from present). **c**, The ratio of values in **b** to those in **a**, giving the change in the risk of a ‘very wet’ winter, arising from human impact on climate.

occurrence of E_2 over Europe from the control ensemble in boreal winter. As would be expected, there is an approximately 2.5% probability of occurrence (consistent with an expected return period of about 40 years). Figure 1b shows the probability of E_2 at the time of CO_2 doubling. Over much of central and northern Europe, the probability has increased to over 12% (consistent with an expected return period of about 8 years), associated with enhanced storm-track activity and 'wetter' storms. By contrast, over parts of the Mediterranean and Northern Africa, the probability of E_2 at the time of CO_2 doubling has decreased.

In Fig. 1c we show the ratio of the probabilities in Fig. 1b and a, giving a measure of the changing risk of E_2 resulting from anthropogenic forcing. Over parts of northern Europe, including much of

the United Kingdom, $r \approx 5$. Defining a 'very wet' winter as one in which E_2 occurs, and with the caveats that the greenhouse ensemble is based on an idealized anthropogenic forcing scenario, and that the CMIP2 ensemble may not necessarily span all model uncertainties, we can give the following assessment: the probability of occurrence of a very wet winter over the UK is estimated to increase by a factor of 5 over the next 50–100 years, due to man's effect on climate.

Corresponding estimates of the changing probability of E_2 over the Asian monsoon region in boreal summer are given in Fig. 2. Note that regions of enhanced probability of E_2 intersect the catchment basin of the Brahmaputra River in particular, but also those of the Ganges and Meghna rivers. Over and above the effects of sea-level rise (not the subject of the present study), this implies an increased risk of flooding in Bangladesh. Figures 1 and 2 extend earlier results⁹ on extremes in daily precipitation to the seasonal timescale and to a multi-model ensemble.

We consider that such probabilistic projections have greater potential value than deterministic projections from either single integrations or from the 'consensus' approach used extensively in the recent IPCC assessment. To show this, a generic 'cost-loss' binary-decision model used to assess the economic utility of ensemble weather forecasts and seasonal predictions^{3,4} was adapted for the climate-change timescale. (We retain the symbols C and L in this adaptation, though their meaning is somewhat different from the original 'cost of protecting against E ' and 'loss due to the occurrence of E '.)

Although such a model is too simple to apply quantitatively to specific realistic decision problems, its underlying formulation deals here with an imagined set of individuals faced with a long-term investment decision: buy either domestic property A (which is sited in a visually attractive location, but which might be prone to flooding), or domestic property B (which is insensitive to weather, but which lies in an area that is likely to be earmarked for high-density development). We assume that all values are measured relative to some inflation-adjusted index, and base the calculations on anticipated climate change over the next 80 years. If no extreme events E_n occur, then the value of A is expected to exceed that of B by an amount C . On the other hand, the value of A will linearly decrease by an amount ΔL every time E_n occurs.

We assume, for simplicity and to minimize sampling effects, that the devaluation ΔL applies if E_n occurs, irrespective of season. For each grid point, let $p(t_{ij})$ denote an estimate of the probability of occurrence of E_n for season i ($i = 1, \dots, 4$) and year j ($j = 1, \dots, 80$). Then the expected devaluation of A due to E_n is $\sum_{j=1}^{80} \sum_{i=1}^4 p(t_{ij}) \Delta L = \bar{p}L$, where $\bar{p}$ is the average probability of occurrence of E_n over the 80-year period, and $L = \Delta L \times 320$. Hence, a rational risk-neutral decision strategy is: buy A if $C > \bar{p}L$, buy B if $C < \bar{p}L$, and toss a coin in the rare case that $C = \bar{p}L$.

Imagine three groups of individuals, each uniformly distributed around the (land points of the) globe. The individuals in group 1 estimate $\bar{p}$ from the probability of occurrence of E_n in the full greenhouse ensemble. By contrast, the individuals in group 2 estimate $\bar{p}$ from the number of occurrences of E_n in a single randomly chosen member of the greenhouse ensemble. (For example, they might rely on a specific national model.) The individuals in group 3 estimate $\bar{p}$ from the number of occurrences of E_n in the greenhouse-ensemble consensus projection.

We estimate the value of the investments made by these three groups, assuming that the true number of occurrences of E_n during the 80-year period is given by a randomly chosen member of the greenhouse ensemble (which is excluded from the climate prediction data set). The value is compared with a lower reference value based on the estimate $\bar{p}$ given by the control ensemble, and an upper reference value based on knowing the truth in advance.

Let M_i denote the expected value of investments for all individuals in subset i at year 80, as aggregated over all land grid points of the

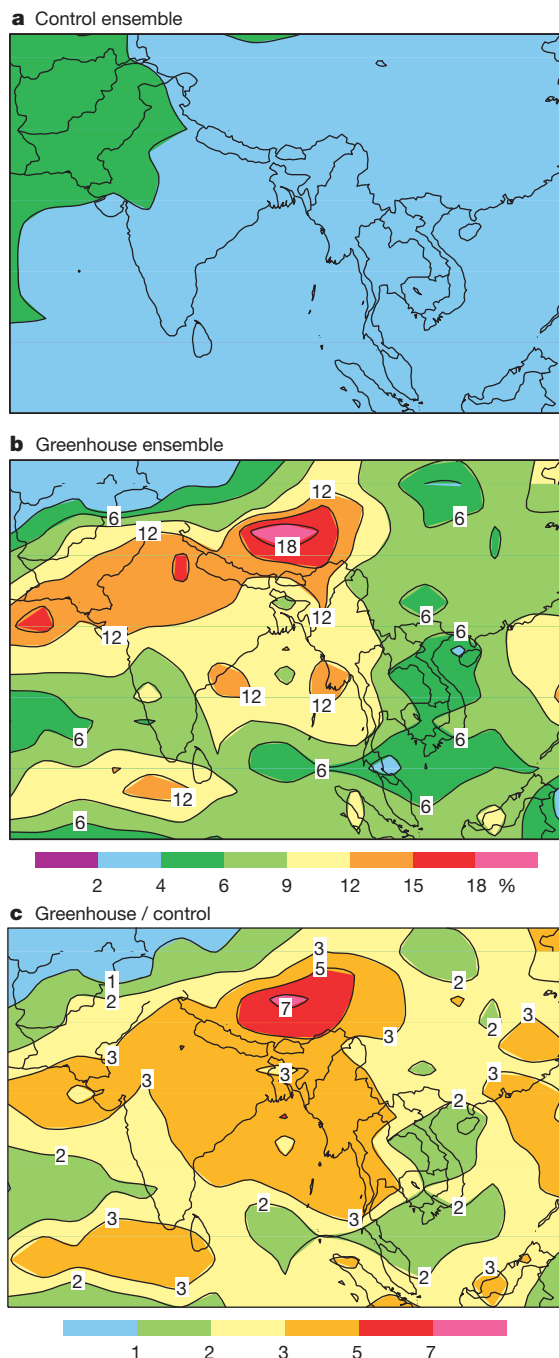


Figure 2 As Fig. 1 but for total boreal summer precipitation in the Asian monsoon region.

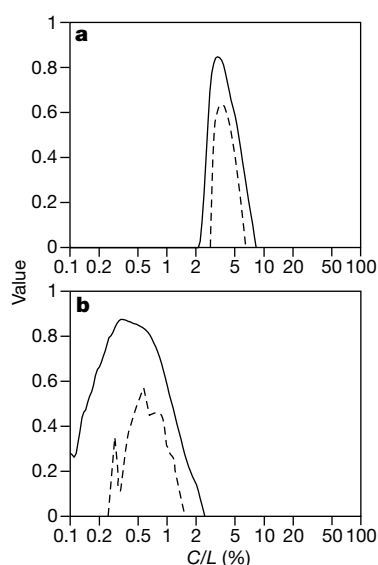


Figure 3 The potential economic value of probabilistic projections of climate change based on a generic binary decision model (see text for details). C and L are possible financial losses, relative to some reference index, associated with two alternative investments. Solid lines, value based on the CMIP2 greenhouse ensemble probability of E_n over the next 80 years. Dashed lines, value based on an estimate of E_n from a single member of the CMIP2 greenhouse ensemble. When value is zero or less, the corresponding estimate of the probability of E_n is no better than an estimate based on the control ensemble. A value of one would correspond to a perfect decision strategy. **a**, $n = 2$; **b**, $n = 3$.

globe. This value is illustrated in terms of the non-dimensional quantity $V_i = (M_i - M_l)/(M_u - M_l)$, where M_l and M_u are the lower and upper reference values respectively. Hence if $V_i = 0$, then the estimate of $\bar{p}$ of subset i is no better than an estimate which assumes that the climate of the twenty-first century will be the same as that of the twentieth century. On the other hand, perfect investment decisions have a value $V_i = 1$. In calculating V_i we average over all choices of verifying model. For V_2 we also average over all choices of forecast model.

Figure 3 shows V_1 and V_2 for $n = 2$ and $n = 3$, as a function of C/L . We do not show V_3 because it is never greater than zero. The reason for this is straightforward: the number of occurrences of extreme climate events is severely underestimated in the consensus projection. This arises because ensemble averaging tends to produce smooth fields which are unrealistically biased towards the climate-mean state.

If C/L is either sufficiently small, or sufficiently large, then accurate ensemble forecasts are clearly not needed to make good investment decisions (buy B or A, respectively). Hence, the greenhouse ensemble only has positive economic value (compared with the baseline value M_l) for a subset of possible C/L . This ensemble tends to be most valuable (compared with M_l) when C/L is comparable with the control ensemble estimate of $\bar{p}$. As $\bar{p}$ is smaller for E_3 than for E_2 , then the greenhouse ensemble provides value for smaller C/L , when decisions are based on the occurrences of E_3 rather than E_2 . It can be seen that the value of the greenhouse ensemble is never less than the value of a single deterministic projection, and that, overall, the difference in value between the ensemble and the single deterministic projection is larger for E_3 than for E_2 . This is because the relative unreliability of the single deterministic projection is larger for E_3 than for E_2 .

In fact, much larger ensembles are needed to provide reliable estimates of the probability of E_3 on a regional basis (for example, specific to the UK or for Bangladesh). Moreover, more research is needed to develop sound methodologies for representing model

uncertainty in climate-prediction ensembles¹⁰. In addition, such extreme-event risk analysis would benefit from an increase in the resolution of present-day climate models (with grid sizes of hundreds of kilometres). For example, it would clearly be of more direct relevance for the decision problem discussed above if E was defined by some specific river bursting its banks during a particular season. However, estimating the future risk of this type of event would require feeding climate model output into a basin-specific hydrological model. In general, climate model grid sizes of the order of ten kilometres would be necessary to simulate precipitation statistics adequately over a typical catchment area (though for rivers with large catchment basins, such as the Brahmaputra, coarser climate model grids may be adequate). An ability to estimate reliably the probability of extreme climate events is an important factor in defining future computational requirements for the climate change problem. □

Received 27 September; accepted 18 December 2001.

1. Houghton, J. T. et al. (eds) *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, 2001).
2. Palmer, T. N. Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys.* **63**, 71–116 (2000).
3. Richardson, D. S. Skill and relative economic value of the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* **126**, 649–668 (2000).
4. Palmer, T. N., Brankovic, C. & Richardson, D. S. A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. *Q. J. R. Meteorol. Soc.* **126**, 2013–2034 (2000).
5. Marsh, T. J. The 2000/01 floods in the UK—a brief overview. *Weather* **56**, 343–345 (2001).
6. Räisänen, J. & Palmer, T. N. A probability and decision-model analysis of a multi-model ensemble of climate change simulations. *J. Clim.* **14**, 3212–3226 (2001).
7. Meehl, G. A., Boer, G. J., Covey, C., Latif, M. & Stouffer, R. J. The coupled model intercomparison project. *Bull. Am. Meteorol. Soc.* **81**, 313–318 (2000).
8. Palmer, T. N. A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrisation in weather and climate prediction models. *Q. J. R. Meteorol. Soc.* **127**, 279–304 (2001).
9. Kharin, V. V. & Zwiers, F. W. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. *J. Clim.* **13**, 3760–3788 (2000).
10. Allen, M. R. Do-it-yourself climate prediction. *Nature* **401**, 642 (1999).

Acknowledgements

We thank M. Blackburn and P. J. Webster for comments on an earlier draft of this Letter.

Increasing risk of great floods in a changing climate

P. C. D. Milly*, R. T. Wetherald†, K. A. Dunne* & T. L. Delworth†

* US Geological Survey, GFDL/NOAA; and † Geophysical Fluid Dynamics Laboratory/NOAA, P.O. Box 308, Princeton, New Jersey 08542, USA

Radiative effects of anthropogenic changes in atmospheric composition are expected to cause climate changes, in particular an intensification of the global water cycle¹ with a consequent increase in flood risk². But the detection of anthropogenically forced changes in flooding is difficult because of the substantial natural variability³; the dependence of streamflow trends on flow regime^{4,5} further complicates the issue. Here we investigate the changes in risk of great floods—that is, floods with discharges exceeding 100-year levels from basins larger than 200,000 km²—using both streamflow measurements and numerical simulations of the anthropogenic climate change associated with greenhouse gases and direct radiative effects of sulphate aerosols⁶. We find that the frequency of great floods increased substantially during the twentieth century. The recent emergence of a statistically

significant positive trend in risk of great floods is consistent with results from the climate model, and the model suggests that the trend will continue.

A focus on large basins facilitates the initiation of quantitative analyses of flood risk with numerical climate models; basins commonly used in trend analyses^{4,5} are too small to be resolved by current models. Estimation of extreme-event risk requires multi-decadal observational records. Here, we consider 29 basins larger than 200,000 km² in area for which discharge observations span at least 30 yr. We analyse annual maximum monthly-mean flows, rather than annual maximum instantaneous flows; these two are strongly correlated in large basins. In contrast with earlier studies^{4,5}, this investigation has a global scope and focuses on extreme events; we analyse the 100-yr flood (that is, the river discharge that has a probability of 0.01 of being exceeded in any given year), which is commonly used in flood-risk assessment for river-basin planning and design of major structures. Choosing such a large-magnitude threshold probably reduces any distortion of our analysis by non-climatic factors such as land-use changes and river development.

For each basin, we fitted observed annual maximum monthly-mean discharges to Pearson's type III distribution⁷ by the method of moments, and determined the 100-yr flood magnitude from the fitted distribution. The 100-yr flood was exceeded 21 times in our observational record of 2066 station-years. Flood events were concentrated in the latter half of the record; half of the observations

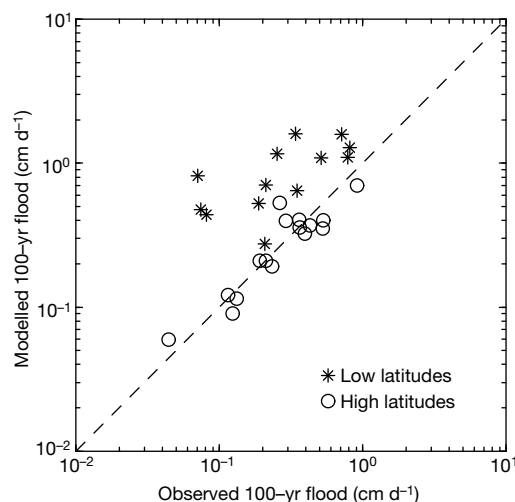


Figure 1 Estimates of 100-yr flood discharges based on model (control experiment years 201–900) and observations (full period of record). Discharges are expressed as monthly-mean discharges per unit basin area. The 16 high-latitude basins are those in North America, Europe, and northern Asia (Fig. 2), and are listed in Table 1. The starting point for basin selection was a previous synthesis and analysis of available streamflow records¹³, in which rivers and gauging sites were chosen to minimize non-climatic sources of non-stationarity. That analysis yielded 34 non-nested basins that satisfy the present constraints on basin area (>200,000 km²) and record length (at least 30 yr). Following more extensive tests for non-climatic non-stationarity in those records, we eliminated four basins on the basis of identified effects of dam construction and one basin on the basis of a change in the gauging site. Collectively, the 2066 station-years of observations on the remaining 29 basins span the 135-yr period 1865–1999; the average record length is 71 yr. For the observations and the model, annual maximum monthly-mean discharges were fitted to Pearson's type III distribution⁷ by the method of moments, and the 100-yr flood was determined from the fitted distribution. To obtain discharge in the climate model, runoff was aggregated over all model cells upstream of the stream gauge of interest and routed through a reservoir whose storage is proportional to its outflow (the modelled river discharge at the gauge site). The constant of proportionality is a residence time whose value was assigned *a priori* on the basis of an analysis of observed basin-mean precipitation and discharge power spectra¹⁴.

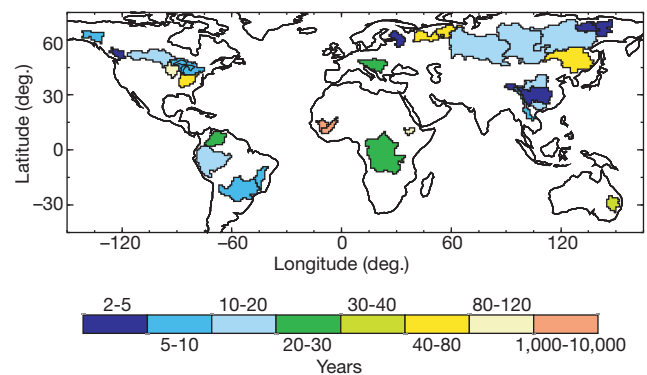


Figure 2 Map showing the gauged drainage areas and flood-risk sensitivities of the 29 river basins in this study. Colour indicates the modelled return period, under idealized quadrupling of atmospheric CO₂ concentrations, of the flood magnitude associated with a 100-yr return period in the control experiment. Although results for low-latitude basins are provided, the poor performance of the model in low latitudes should be kept in mind.

were made after 1953, and 16 of the flood events occurred after 1953. Under the assumption that flood events were independent outcomes of a stationary process, we used binomial probability theory to determine a probability of 1.3% of having 16 or more of 21 events during the second part of the record. For observations from an extratropical subset of the basins (see below), the corresponding probability is 3.5%, for 7 out of 8 flood events in the second half of the record. Supplementary analyses for shorter return periods (2–50 yr) did not reveal significant trends, but 200-yr flood frequency increased significantly.

We now refine the simple analysis above to address certain shortcomings: the assumption of independence among flood events, the use of a crude index of the trend based on simple bisection of the historical sequence, and the presence of sampling errors in our estimates of 100-yr flood magnitudes. We first introduce a more robust measure, *Z*, of the flood-frequency trend; *Z* is the slope of the least-squares linear relation between annual flood frequency (number of flood events divided by number of operating stations) and time, with annual data values weighted by number of operating stations. To estimate the probability density function of *Z* under constant climate, we used output from a 900-yr

Table 1 Effect on river discharge of quadrupling atmospheric CO₂

River	Station	dq (%)	dQ (%)	P (%)
Yukon	Eagle, USA	54	15	11
Nelson	Bladder Rapids, Canada	76	36	8.8
Fraser	Hope, Canada	44	17	27
St Lawrence	Cornwall, Canada	20	15	12
Mississippi	Keokuk, USA	14	–1	0.90
Ohio	Metropolis, USA	–12	9	2.3
Danube	Orsova, Romania	–3	16	4.6
Neva	Novosaratovka, Russia	36	22	31
N. Dvina	Ust-Pinega, Russia	39	2	1.4
Pechora	Ust-Tsilma, Russia	25	5	2.2
Ob'	Salekhard, Russia	41	14	7.7
Yenisei	Igarka, Russia	23	13	5.9
Lena	Kusur, Russia	25	25	9.9
Yana	Dzanghky, Russia	42	43	32
Indigirka	Vorontsovo, Russia	42	30	46
Amur	Khabarovsk, Russia	4	10	2.4

Changes in discharge of extratropical rivers associated with idealized quadrupling of atmospheric CO₂ concentration in the model. dq is the relative change in annual mean discharge, dQ is the relative change in 100-yr annual maximum monthly discharge, and P is the annual probability of 100-yr flood (defined with respect to the control experiment), after quadrupling of atmospheric CO₂. We used the 100 yr of model output that begins 60 yr after stabilization of CO₂ concentration at the quadrupled level. Post-quadrupling distributions of annual maximum monthly flows were fitted to Pearson's type III distribution, which was then used to determine the probability of the control 100-yr flood.

'control' (constant radiative forcing) experiment with a coupled ocean-atmosphere-land model⁶. The model simulates well 100-yr flood thresholds (and annual discharge statistics) for basins far outside the tropics, but systematically overestimates flood magnitudes in the lower latitudes (Fig. 1); accordingly, we performed significance analyses both for the full set of basins and for the subset of 16 higher-latitude ('extratropical') basins, and we confined much of the subsequent analysis to the extratropical domain. For the historical record, $Z = 1.99 \times 10^{-4} \text{ yr}^{-1}$ with all basins ($1.41 \times 10^{-4} \text{ yr}^{-1}$ extratropical).

To evaluate the significance of these values of Z , we extracted from the control experiment 500 overlapping (hence, non-independent) 135-yr sequences (years 177–311, 178–312, ..., 676–810) of flows, mapped each to the time period 1865–1999, and sampled these sequences for the river-specific periods of observations in the historical record. For each sequence, we estimated the 100-yr flood level for each basin (thereby simulating the sampling error inherent in the observational analysis), determined flood occurrences, and calculated Z . The observed value of Z was exceeded in none of the 500 sequences when all basins were considered, and was exceeded in 3.2% of the sequences when only extratropical basins were considered. (Overall trends in flood frequency over the analysed part of the control experiment were negligible.) Thus, the model-based significance analysis, which implicitly uses the space-time correlation structure of floods in the model, essentially confirms and reinforces the simpler binomial analysis.

The apparent increase in flood risk might be associated with radiatively forced climate change. To assess flood-risk sensitivity to radiative forcing, we used a 300-yr 'idealized CO₂ quadrupling' experiment with a 1%-per-year growth (for 140 yr) of atmospheric

CO₂ concentration from the control level to a stable, quadrupled level⁸ (maintained for 160 yr). Modelled changes in annual mean discharge (relative to the control experiment) provide a measure of intensification of the water cycle as a result of idealized CO₂ quadrupling (Table 1). For the extratropical basins, these range from –12% to +76%, with a median value of +30%. Relative changes in the 100-yr monthly maximum discharge generally are smaller and less variable, with a median of +15%. In all but one of the basins, the control 100-yr flood is exceeded more frequently as a result of idealized CO₂ quadrupling (Fig. 2). The probability of exceeding this control flood changes by a factor that ranges from 0.90 to 46; in half of the basins, the factor exceeds 8 (implying a decrease in return period from 100 yr to shorter than 12.5 yr).

Given the substantial modelled sensitivity of flood risk to radiative forcing, we framed the hypothesis that historical changes in flood risk. We examined the detectability of flood-risk change in five transient 'scenario' climate experiments (225 yr, 1865–2089) that shared common estimates of historical and projected future changes in radiative forcing by greenhouse gases and direct effects of sulphate aerosols, each with a distinct initial condition⁶. These experiments show an increase in extratropical flood frequency that generally is apparent early in the twenty-first century (Fig. 3). Thereafter, the flood rate is 2 to 8 times greater than its value during the historical period of observations. Values of extratropical Z were computed for each scenario with exactly the same gauging schedule as in the observations. Four of the experiments (all except scenario 4) had positive values of Z ; the largest of these ($1.32 \times 10^{-4} \text{ yr}^{-1}$, in scenario 3) was slightly smaller than the observed value.

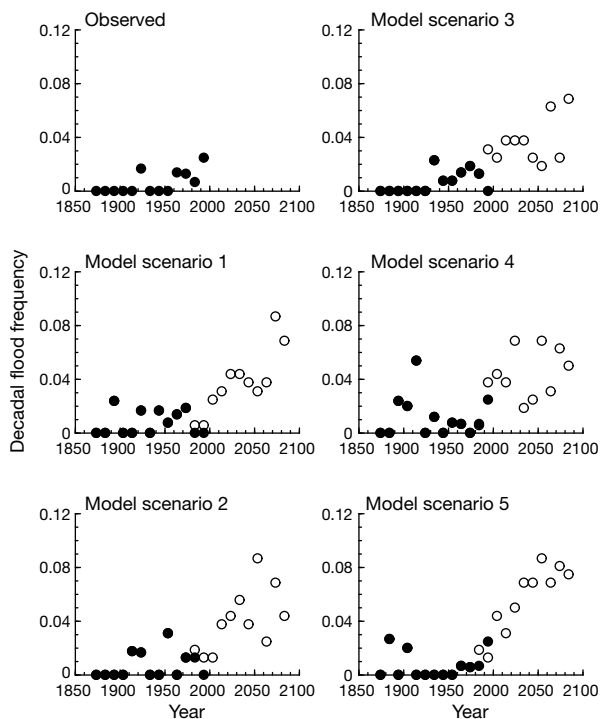


Figure 3 Decadal extratropical flood frequencies for observations and for scenario experiments. Flood frequency is the number of events exceeding the 100-yr discharge, divided by the number of station-years of observations. In plots of model outputs, filled circles are obtained with each station starting and ending as in the historical record; open circles are obtained with all stations continuing operation once begun. The symbols coincide before the 1980s, because no station record ends until 1986. Distinct 100-yr flood levels were defined for each of the six data sets with respect to the historical schedule of observations over that data set.

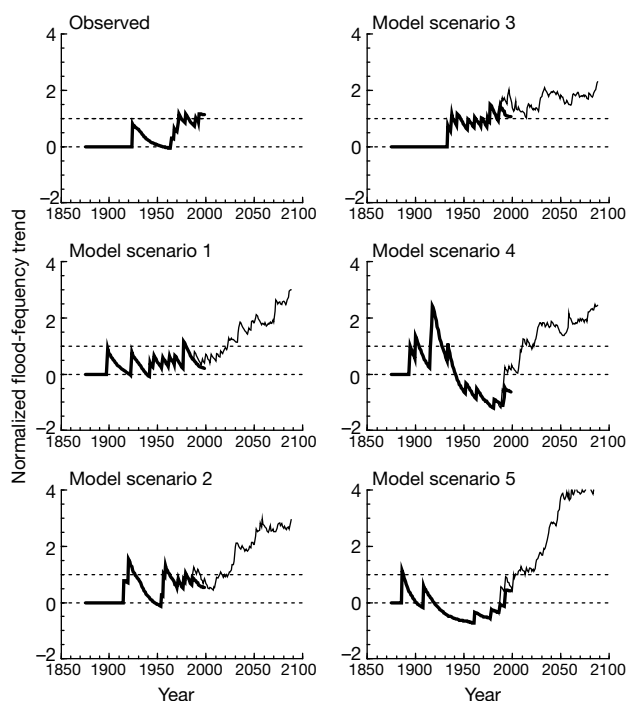


Figure 4 Normalized trend $Z(t)/Z_{95}(t)$ of extratropical flood frequency for observations and for scenario experiments. $Z_{95}(t)$ is the value of the flood-frequency trend statistic $Z(t)$ that equals or exceeds the values for 475 (95%) of the 500 control-experiment segments (see text). Normalized trends greater than 1 imply significance with respect to this measure. Results indicated by bold curves are obtained with each station starting and ending as in the historical record; results shown by light curves are obtained with all stations continuing operation once begun. These curves are identical before 1986.

Because trends and their statistical significance are random functions of time, an analysis of the time variation of detectability of flood-frequency change may be informative. Therefore, we generalized Z to $Z(t)$, where t is the hypothetical last year of available records, and evaluated $Z(t)$ for the observations and the scenario experiments (Fig. 4). The observed flood trend $Z(t)$ has been significantly (at 95% level, as evaluated from the control experiment) different from zero continuously since the flooding of the upper Mississippi River in 1993 and intermittently since 1972. Uninterrupted periods of statistically significant flood-frequency trends in scenarios 1 to 5 begin in years 2023, 2023, 1986, 2021 and 2006, respectively, and premonitory multidecadal periods of intermittent significance begin much earlier in scenarios 2 and 3 (1956 and 1937, respectively). Thus, the recent history of the observed trend index is generally consistent with the range of results from the scenario experiments.

Our detection of an increase in great-flood frequency and its attribution to radiatively induced climate change are tentative. The frequency of floods having return periods shorter than 100 yr did not increase significantly. Potentially significant effects of measurement non-stationarity are not easily assessed. The forced signal and unforced variability in the model contain errors of unknown magnitude. Absent from the model are forcings such as solar variability, volcanic activity, land-cover change⁹, and water-resource development¹⁰, and potential biospheric feedbacks such as CO₂-induced stomatal closure¹¹ and water-stress-induced root extension¹². Especially evident from our study are the needs for improvements in simulation of tropical hydroclimate and continued commitment to stream-gauging programmes worldwide. □

Received 2 October; accepted 17 December 2001.

1. Cubasch, U. in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. et al.) Ch. 9 (Cambridge Univ. Press, Cambridge, 2001).
2. White, K. S. et al. Technical Summary in *Climate Change 2001: Impacts, Adaptation and Vulnerability* (eds McCarthy, J. J. et al.) 19–73 (Cambridge Univ. Press, Cambridge, 2001).
3. McCabe, G. J. Jr & Wolock, D. M. Climate change and the detection of trends in annual runoff. *Clim. Res.* **8**, 129–134 (1997).
4. Lins, H. F. & Slack, J. R. Streamflow trends in the United States. *Geophys. Res. Lett.* **26**, 227–230 (1999).
5. Groisman, P. Ya., Knight, R. W. & Karl, T. R. Heavy precipitation and high streamflow in the contiguous United States: Trends in the twentieth century. *Bull. Am. Meteorol. Soc.* **82**, 219–246 (2001).
6. Knutson, T. R., Delworth, T. L., Dixon, K. W. & Stouffer, R. J. Model assessment of regional surface temperature trends (1947–1997). *J. Geophys. Res.* **104**, 30981–30996 (1999).
7. Pearson, K. *Tables for Statisticians and Biometricians* 3rd edn (Cambridge Univ. Press, Cambridge, 1930).
8. Delworth, T. L. et al. Simulation of climate variability and change by the GFDL R30 coupled climate model. *Clim. Dyn.* (submitted).
9. Chase, T. N., Pielke, R. A. Sr, Kittel, T. G. F., Nemani, R. R. & Running, S. W. Simulated impacts of historical land cover changes on global climate in northern winter. *Clim. Dyn.* **16**, 93–105 (2000).
10. Milly, P. C. D. & Dunne, K. A. Trends in evaporation and surface cooling in the Mississippi River basin. *Geophys. Res. Lett.* **28**, 1219–1222 (2001).
11. Sellers, P. J. et al. Comparison of radiative and physiological effects of doubled atmospheric CO₂ on climate. *Science* **271**, 1402–1406 (1996).
12. Milly, P. C. D. Sensitivity of greenhouse summer dryness to changes in plant rooting characteristics. *Geophys. Res. Lett.* **24**, 269–271 (1997).
13. Milly, P. C. D. & Dunne, K. A. Macroscale water fluxes: 1. Quantifying errors in the estimation of basin-mean precipitation. *Wat. Resour. Res.* (submitted).
14. Milly, P. C. D. & Wetherald, R. T. Macroscale water fluxes: 3. Effects of land processes on variability of monthly river discharge. *Wat. Resour. Res.* (submitted).

Acknowledgements

Streamflow data were collected and/or provided to us by the Global Runoff Data Centre (Germany), the US Geological Survey, Environment Canada, EarthInfo, Manaus Harbor Ltd/Portobras and Departamento Nacional de Aguas e Energia Elétrica (Brazil), National Institute of Meteorology and Hydrology (Romania), J. Cordova, N. Garcia and J. Richey. We thank K. L. Findell, D. P. Lettenmaier and R. J. Stouffer for comments and suggestions.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.C.D.M. (e-mail: cmilly@usgs.gov).

Antarctic climate cooling and terrestrial ecosystem response

Peter T. Doran*, John C. Priscu†, W. Berry Lyons‡, John E. Walsh§, Andrew G. Fountain||, Diane M. McKnight¶, Daryl L. Moorhead#, Ross A. Virginia☆, Diana H. Wall**, Gary D. Clow††, Christian H. Fritsen‡‡, Christopher P. McKay§§ & Andrew N. Parsons**

* Department of Earth and Environmental Sciences, University of Illinois at Chicago, 845 West Taylor Street, Chicago, Illinois 60607, USA

† Land Resources and Environmental Sciences, 334 Leon Johnson Hall, Montana State University, Bozeman, Montana 59717, USA

‡ Byrd Polar Research Center, Ohio State University, 1090 Carmack Road, Scott Hall, Columbus, Ohio 43210, USA

§ Department of Atmospheric Sciences, University of Illinois, 105 South Gregory Street, Urbana, Illinois 61801, USA

|| Department of Geology, Portland State University, Portland, Oregon 97207, USA

¶ Institute of Arctic and Alpine Research, 1560 30th Street, Campus Box 450, Boulder, Colorado 80309, USA

Department of Earth, Ecological and Environmental Sciences, 2801 W. Bancroft Street, University of Toledo, Toledo, Ohio 43606, USA

☆ Environmental Studies Program, Dartmouth College, 6182 Steele Hall, Hanover, New Hampshire 03755, USA

** Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, Colorado 80523, USA

†† USGS—Climate Program, Box 25046, MS 980, Denver Federal Center, Denver, Colorado 80225, USA

‡‡ Division of Earth and Ecosystem Sciences, Desert Research Institute, 2215 Raggio Parkway, Reno, Nevada 89512, USA

§§ Space Science Division, NASA Ames Research Center, Moffet Field, California 94035, USA

The average air temperature at the Earth's surface has increased by 0.06 °C per decade during the 20th century¹, and by 0.19 °C per decade from 1979 to 1998². Climate models generally predict amplified warming in polar regions^{3,4}, as observed in Antarctica's peninsula region over the second half of the 20th century^{5–9}. Although previous reports suggest slight recent continental warming^{9,10}, our spatial analysis of Antarctic meteorological data demonstrates a net cooling on the Antarctic continent between 1966 and 2000, particularly during summer and autumn. The McMurdo Dry Valleys have cooled by 0.7 °C per decade between 1986 and 2000, with similar pronounced seasonal trends. Summer cooling is particularly important to Antarctic terrestrial ecosystems that are poised at the interface of ice and water. Here we present data from the dry valleys representing evidence of rapid terrestrial ecosystem response to climate cooling in Antarctica, including decreased primary productivity of lakes (6–9% per year) and declining numbers of soil invertebrates (more than 10% per year). Continental Antarctic cooling, especially the seasonality of cooling, poses challenges to models of climate and ecosystem change.

Terrestrial ecosystem research in the Antarctic is restricted to a few ice-free areas of the coast, including the McMurdo Dry Valleys (77–78° S, 160–164° E). The dry valleys region is the largest ice-free area on the Antarctic continent. It is a cold desert, comprising a mosaic of perennially ice-covered lakes, ephemeral streams, arid soils, exposed bedrock, and alpine glaciers. Published historical weather observations in the dry valleys are limited^{11–14}. Biological activity is microbially dominated and diversity is low. The largest animals are soil invertebrates, of which soil nematodes are the most widely distributed¹⁵.

Our 14-year, continuous automatic weather station record from the shore of Lake Hoare reveals that seasonally averaged surface air temperature has decreased by 0.7 °C per decade ($P = 0.21$) from 1986 to 1999 (Fig. 1a). The temperature decrease is most pronounced in

the summer (December–February = 1.2°C per decade, $P = 0.02$) and autumn (March–May = 2.0°C per decade, $P = 0.11$). Winter (June–August) and spring (September–November) show smaller temperature increases (0.6°C and 0.1°C per decade, $P = 0.62$ and 0.95 , respectively). The dry valley cooling, and its seasonal pattern

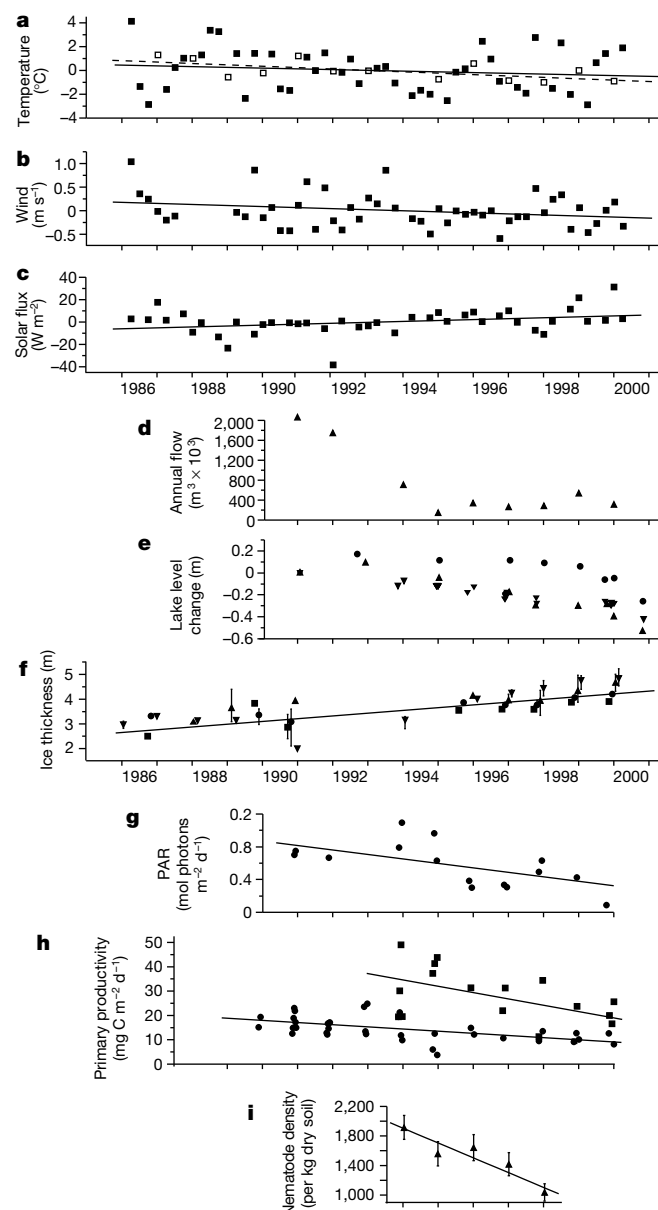


Figure 1 Meteorological and ecosystem changes in the McMurdo Dry Valleys, 1986–2000. **a**, Seasonally averaged air temperature over time at Lake Hoare station (full data set) with summer values highlighted (open squares). Trend lines are annual (solid line) and summer only (dashed lines). **b**, Seasonally averaged wind speed over time at Lake Hoare station. **c**, Seasonally averaged solar flux at Lake Hoare station (excluding winter, June–August). **d**, Total annual stream flow from eight streams in the Lake Fryxell basin. **e**, Lake level change in Lake Hoare (inverted triangles), Lake Fryxell (triangles), and Lake Bonney (circles). **f**, Lake ice thickness of west lobe Lake Bonney (squares), east lobe Lake Bonney (circles), Lake Fryxell (triangles), and Lake Hoare (inverted triangles). Ice thickness is measured from the water level to the bottom of the ice in holes drilled through the ice. The vertical bars indicate the range of measurements within a season. **g**, Mean monthly (November and December only) photosynthetically active radiation (PAR) 10 m below the surface of the ice in the east lobe of Lake Bonney. **h**, Depth-integrated primary productivity during November and December in east (circles and lower trend line) and west (squares and upper trend line) lobes of Lake Bonney. **i**, Total number of soil nematodes over time in experimental plots on the south shore of Lake Hoare.

(that is, dominated by summer and autumn), reflects longer term continental Antarctic cooling between 1966 and 2000 (Fig. 2 and Table 1). Owing to the exclusion of dry valley records in Fig. 2, compatibility with the dry valley data increases the validity of the analysis. Moreover, Fig. 2 is consistent with maps of individual station trends during 1976–2000 presented in the Intergovernmental Panel on Climate Change (IPCC) report¹. We focus here on the Lake Hoare record because it is the longest dry valley record, but seven other dry valley floor stations show similar trends¹⁶.

The seasonally averaged wind speed decreased by 0.23 m s^{-1} per decade ($P = 0.07$) at Lake Hoare from 1986 to 1999 (Fig. 1b), and is significantly correlated with the seasonally averaged temperature decrease ($P < 0.01$). Furthermore, both easterly on-shore coastal (dominant in the summer) and westerly katabatic (dominant in winter, spring and autumn) wind speeds are significantly correlated ($P < 0.01$) with temperature. Annual temperatures at individual dry valley sites are strongly controlled by exposure to wind; the dry adiabatic lapse rate and distance to the coast are of secondary importance¹⁶. Our results suggest that estimating long-term temperature change in coastal Antarctica requires an understanding of the synoptic controls on surface wind variability, which at present are incompletely understood^{17,18}.

Seasonally averaged (excluding June–August) solar radiation has increased from 1986 to 1999 by 8.1 W m^{-2} per decade ($P = 0.05$; Fig. 1c). Radiation during non-winter months decreased with increasing wind speed during this period ($P = 0.08$). The inverse relationship between wind speed and radiation is highly significant for spring and autumn ($P < 0.01$). Radiation decreases significantly with easterly wind speed during the summer ($P = 0.02$), and with westerly wind speed during the spring ($P = 0.03$). Observers in the field routinely noted cloudiness during high wind events. Together, these results indicate that increased solar radiation in the dry valleys is related to decreased wind and associated cloudiness over time.

Changes in dry valley moisture indices (relative humidity and precipitation) are inconclusive because of measurement uncertainties. Snow accumulation (precipitation minus evaporation) on two local valley glaciers showed no clear trend between the summers of 1993–94 and 1999–2000. We infer from the increased clear-sky conditions that cloudiness decreased from 1986 to 1999. Soil moisture decreased from 2.2% (by weight) to 1.4% between 1993 and 1999 in an elevational transect in Taylor Valley.

The McMurdo Dry Valley environment has long been viewed as sensitive to low amplitude climatic shifts^{19–21}. Local hydrology is dependent on small changes in summer temperature and solar radiation, which can melt glacier ice and provide liquid runoff to soils, streams and lakes. From 1969 to 2000, all discharge observations were made in dry valley streams between 10 November and 24 March, but typically most flow occurs in December and January. All streams are fed largely by glacial melt, with minor inputs coming from seasonal snow banks. Storage of stream water occurs in hyporheic zones: moist soil areas adjacent to and beneath the streams²². The discharge from the eight principal inflow streams in the Lake Fryxell basin since 1990–91 (except 1992–93 when field measurements were not made) decreased nonlinearly by an average rate of $1.8 \times 10^5\text{ m}^3\text{ yr}^{-1}$ (Fig. 1d). Total annual stream flow in these streams increased by $41,000\text{ m}^3$ per degree-day above freezing at the

Table 1 Proportions of Antarctica warming and cooling (1966–2000)

Period	Antarctica	Antarctica without the Antarctic Peninsula
Annual	+41.4%, –58.3%	+33.8%, –65.9%
Winter (June–Aug.)	+62.5%, –37.3%	+56.3%, –43.4%
Spring (Sept.–Nov.)	+54.1%, –45.7%	+49.4%, –50.4%
Summer (Dec.–Feb.)	+31.7%, –67.4%	+22.8%, –76.3%
Autumn (Mar.–May)	+12.6%, –87.4%	+0.3%, –99.7%

Plus signs indicate the proportions warming; minus signs indicate the proportions cooling. The Antarctic Peninsula is defined as the area north of 80°S and east of 80°W .

Lake Fryxell meteorological station ($P < 0.01$).

Lake levels rose at an average rate of 16 cm yr^{-1} between 1903 and approximately 1990, and lake ice thinned before 1986^{19,21}. Our data show that lake levels receded (Fig. 1e) in response to cooler summers and decreased stream flow since 1990. Cooler, quiescent conditions in summer reduce sublimation loss from lakes, but not enough to compensate for the decreased stream flow. The thickness of lake ice has increased since 1986 by an average of 11 cm yr^{-1} ($P < 0.01$) in response to the lower temperatures (Fig. 1f).

We believe that climate cooling has significantly impacted ecosystem properties in the McMurdo Dry Valleys. The climate-induced increase in lake ice thickness has reduced underwater irradiance during November–December in the east lobe of Lake Bonney by $0.055 \text{ mol photons m}^{-2} \text{ d}^{-1}$ ($P = 0.01$) since 1990 (Fig. 1g). Because phytoplankton primary production in the dry valley lakes is limited by light²³, we suggest that this decrease in irradiance has affected the rate of primary production in the lakes (Fig. 1h). Depth-integrated primary production in the east and west lobes of Lake Bonney has decreased by 0.88 ($P < 0.01$) and 2.6 ($P = 0.03$) $\text{mg C m}^{-2} \text{ d}^{-1}$ annually, amounting to a 6% and 9% decrease per year, respectively. Recent data on the carbon biogeochemistry of Lake Bonney show that contemporary photosynthesis to respiration ratios are less than unity²⁴. The inferred climate-induced reduction in primary production will exacerbate this situation producing a system that may act as a CO_2 source and eventually become depleted in organic carbon stores. Reduced nutrient loading associated with decreasing stream flows is not the cause of the noted reduction in primary production, as a large portion of the nutrient supply for phytoplankton growth arises from internal vertical diffusion²⁴.

Soil invertebrate communities showed changes in diversity and abundance from 1993 to 1998. The abundance of tardigrades and nematodes, including the dominant nematode species *Scottinema lindsayae*, declined in an elevational transect²⁹, and across all treatments (moisture, temperature, and carbon) in a climate

manipulation experiment by 200 individuals ($>10\%$) per year ($P = 0.01$, Fig. 1i). Given the low diversity and long generation times of these invertebrates, these declines in population represent important shifts in the diversity, life cycles, trophic relationships and functioning of dry valley soils²⁵.

We have shown a 14-yr Antarctic dry valley meteorological record and 35-yr continental temperature compilation that indicate an annual cooling trend during recent time over much of the continent of Antarctica, outside of the peninsula. Although other studies (for example, ref. 10) have cited a trend of continental warming in Antarctica, the trends are sensitive to the period analysed and to the distribution of stations. The warming reported in ref. 10 occurs almost entirely from 1958 to 1978, but not thereafter. As the shorter term dry valley data are for the period subsequent to the primary warming, the trends deduced in the different studies are not incompatible. Moreover, the large-scale cooling reported here results from an approach designed to avoid over-weighting of station-dense regions (for example, the peninsula) in the evaluation of overall trends. In the dry valleys, the cooling trend is significantly correlated with decreased winds and increased clear-sky conditions over the period of record. These changes are indicative of the strong influence that winds (mostly onshore during the summer and katabatic during other seasons) have on the dry valley climate. We propose that prolonged summer cooling will diminish aquatic and soil biological assemblages throughout the valleys, and possibly in other terrestrial Antarctic ecosystems. Winter temperatures are well below the threshold for liquid water production and can fluctuate significantly with minimal direct hydrological or ecological impact. Summer temperatures are the critical driver of Antarctic terrestrial ecosystems, and our data are the first, to our knowledge, to highlight the cascade of ecological consequences that result from the recent summer cooling. □

Methods

Dry valley ecosystem parameters

All dry valley meteorological data were collected using Campbell Scientific data loggers. The network consists of four stations in Taylor Valley, two in Wright Valley, and one in Victoria Valley. Precise station locations can be found in ref. 26. Four other stations on glacier surfaces in Taylor Valley are not discussed in this paper. Air temperature was collected at 3 m from the ground using a fenwall-type thermistor in a shielded Campbell Scientific model 207 probe. We calculated all temperature data from raw voltages using a Steinhart–Hart equation. Wind speed was measured at 3 m above the ground using a Met One model 014A wind speed sensor and model 024A wind direction sensor, up to 1993, and an R.M. Young model 05103 wind speed and direction sensor thereafter. Since 1993, we replaced all wind monitors once for recalibration. Solar flux was measured using Li-Cor model LI-200 pyranometers, which have a maximum stated uncertainty of $\pm 5\%$, and were recalibrated against an Eppley pyranometer every 2 yr. These pyranometers do not measure the ultraviolet range absorbed by ozone. Relative humidity was measured at all stations using Phys-Chem humidity transducers in Campbell Scientific 007 probes. Calibration drift in these transducers is significant, thereby possibly obscuring any long-term trends. We measured stream flow using pressure transducers in flumes. We calibrated rating curves frequently during the melt season.

Soil moisture was determined gravimetrically (48 h at 105°C) from 50-g soil samples that were collected in polyethylene bags in the field. On each sampling date, soil moisture was determined for 51 soil samples collected from $10 \times 10 \text{ m}$ grids established at 83, 121 and 188 m elevation near the south shore of Lake Hoare, Taylor Valley. See ref. 27 for further details on the transect.

Irradiance in the water column of the Taylor Valley lakes was measured with a Li-Cor model 193 spherical quantum (400–700 nm) sensor moored 10 m beneath the surface of the permanent lake ice. The data were logged with a Campbell 21x data logger at 20-min intervals throughout the year. Primary productivity in lakes was measured using the ^{14}C method, outlined in ref. 22, during 24-h *in situ* incubations.

For nematode analysis, soil samples were collected with pre-sterilized plastic scoops and placed in sterile polyethylene Whirl-Pak bags. All soils were transported in insulated coolers to the McMurdo station laboratory facilities, where they were immediately stored at 4°C . Nematodes, tardigrades, and rotifers were extracted from the soils within 48 h using standard sugar centrifugation procedures, modified to keep the soils and all extraction materials at a constant cold temperature. Extracted nematodes were identified to genus level. All nematode counts were adjusted for soil moisture to give number of nematodes per kilogram of dry soil.

Continental temperature trends

Continental temperature trend maps were computed from the gridded University of East

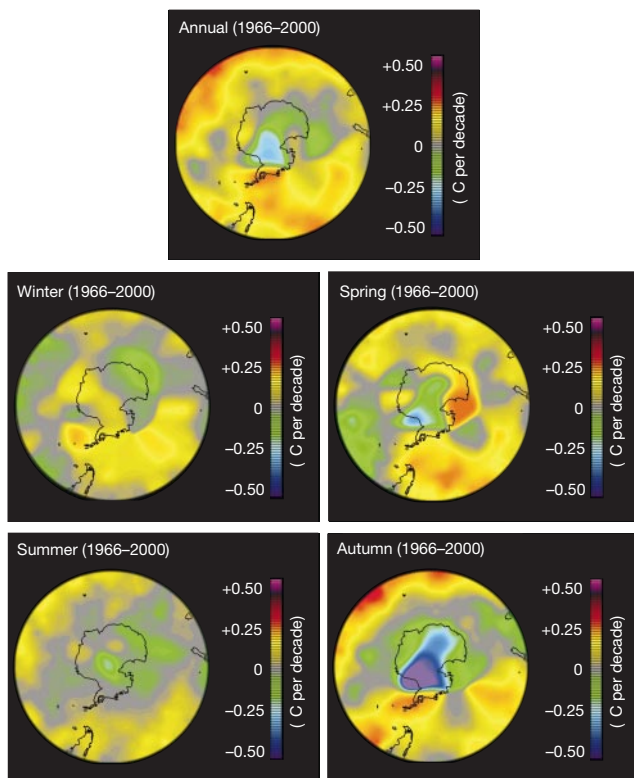


Figure 2 Annual and seasonal Antarctic surface temperature trends ($^\circ\text{C per decade}$) between 1966 and 2000 calculated from the University of East Anglia Climate Research Unit's $5^\circ \times 5^\circ$ HadCRUT data set²⁸.

Anglia HadCRUT temperature data set, based on land station and ship reports²⁸. The trends for each 5° × 5° grid cell were evaluated by a least-squares fit for the period 1965–2000. The gridded trend values were then smoothed spatially using a Cressman analysis, which effectively determines a pixel value as a weighted sum of contributions from surrounding grid points for which data are available. Weights vary as the inverse fourth power of the distance from the pixel in question. The radius of influence is 500 pixels, or approximately one-quarter the maximum width of the image.

Received 28 September; accepted 4 December 2001.

Published online 13 January 2002, DOI 10.1038/nature710.

- Houghton, J. T. *et al.* (eds) *Climate Change 2001: The Scientific Basis. Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, Cambridge, 2001).
- National Research Council *Reconciling Observations of Global Temperature Change* (National Academy Press, Washington DC, 2000).
- Chen, C. T. A. & Drake, E. T. Carbon dioxide increase in the atmosphere and oceans and possible effects on climate. *Annu. Rev. Earth Planet. Sci.* **14**, 201–235 (1986).
- Cattle, H. & Crossley, J. Modeling arctic climate change. *Phil. Trans. R. Soc. Lond. A* **352**, 201–213 (1995).
- Weller, G. Regional impacts of climate change in the Arctic and Antarctic. *Ann. Glaciol.* **27**, 543–552 (1998).
- Vaughan, D. G. & Doake, C. S. M. Recent atmospheric warming and retreat of ice shelves on the Antarctic Peninsula. *Nature* **379**, 328–331 (1996).
- Comiso, J. C. Variability and trends in Antarctic surface temperatures from in situ and satellite infrared measurements. *J. Clim.* **13**, 1674–1696 (2000).
- Smith, R. C. *et al.* Marine ecosystem sensitivity to climate change. *BioScience* **49**, 393–404 (1999).
- Vaughan, D. G. *et al.* Devil in the detail. *Science* **293**, 1777–1779 (2001).
- Jacka, T. H. & Budd, W. F. Detection of temperature and sea-ice-extent changes in the Antarctic and Southern Ocean 1949–96. *Ann. Glaciol.* **27**, 553–559 (1998).
- Riordan, A. J. in *Climate of the Arctic* (eds Weller, G. & Bowling, S. A.) 268–275 (Geophysical Institute, University of Alaska, Fairbanks, 1973).
- Bromley, A. M. *Weather Observations, Wright Valley, Antarctica* Information Publication no. 11 (New Zealand Meteorological Service, Wellington, 1985).
- Clow, G. D., McKay, C. P., Simmons, G. M. Jr & Wharton, R. A. Jr Climatological observations and predicted sublimation rates at Lake Hoare. *Antarct. J. Clim.* **1**, 715–728 (1988).
- McKay, C. P., Nienow, J. A., Meyer, M. A. & Friedmann, E. I. in *Antarctic Meteorology and Climatology: Studies Based on Automatic Weather Stations* Antarctic Research Series 61 (eds Bromwich, D. H. & Stearns, C. R.) 201–207 (American Geophysical Union, Washington DC, 1993).
- Freckman, D. W. & Virginia, R. A. Low-diversity Antarctic soil nematode communities: distribution and response to disturbance. *Ecology* **78**, 363–369 (1997).
- Doran, P. T. *et al.* Climate observations (1986–2000) from the McMurdo Dry Valleys, Antarctica. *J. Clim.* (submitted).
- Bromwich, D. H. & Parish, T. R. (eds) *Antarctica: Barometer of Climate Change* Report of the National Science Foundation Antarctic Meteorology Working Group (National Science Foundation, Arlington, Virginia, 1998).
- Parish, T. R. & Cassano, J. J. Forcing of the wintertime Antarctic boundary layer winds from the NCEP-NCAR global reanalysis. *J. Appl. Meteorol.* **40**, 810–821 (2001).
- Wharton, R. A. *et al.* Changes in ice cover thickness and lake level of Lake Hoare, Antarctica—implications for local climatic change. *J. Geophys. Res.* **97**, 3503–3513 (1993).
- Fountain, A. G. *et al.* Physical controls on the Taylor Valley ecosystem, Antarctica. *BioScience* **49**, 961–971 (1999).
- Chinn, T. J. in *Physical and Biogeochemical Processes in Antarctic Lakes* Antarctic Research Series 59 (eds Green, W. J. & Friedmann, E. I.) 1–51 (American Geophysical Union, Washington DC, 1993).
- McKnight, D. M. *et al.* Dry valley streams in Antarctica: ecosystems waiting for water. *BioScience* **49**, 985–995 (1999).
- Priscu, J. C. *et al.* Carbon transformations in a perennially ice-covered Antarctic lake. *BioScience* **49**, 997–1008 (1999).
- Priscu, J. C. Phytoplankton nutrient deficiency in lakes of the McMurdo Dry Valleys, Antarctica. *Freshwat. Biol.* **34**, 215–227 (1995).
- Virginia, R. A. & Wall, D. H. How soils structure communities in the Antarctic dry valleys. *BioScience* **49**, 973–983 (1999).
- Doran, P. T., Dana, G., Hastings, J. T. & Wharton, R. A. The McMurdo LTER automatic weather network (LAWN). *Antarct. J. US* **30**, 276–280 (1995).
- Powers, L. E., Ho, M., Freckman, D. W. & Virginia, R. A. Distribution, community structure, and microhabitats of soil invertebrates along an elevational gradient in Taylor Valley, Antarctica. *Arct. Alpine Res.* **30**, 133–141 (1998).
- Jones, P. D. *et al.* Surface air temperature and its changes over the past 150 years. *Rev. Geophys.* **37**, 173–199 (1999).
- Porazinska, D. L., Wall, D. H. & Virginia, R. A. Spatial and temporal variation in nematode populations over a six-year period in the McMurdo Dry Valleys, Antarctica. *Arctic Antarct. Alp. Res.* (in press).

Acknowledgements

We thank the personnel associated with the McMurdo Long Term Ecological Research site who contributed to the collection of data. T. Chinn provided the three earliest data points on the lake level plot. W. Chapman assisted with the compilation of the continental figures. This work was supported by the National Science Foundation's Office of Polar Programs, the United States Geological Survey, and the NASA Exobiology Program.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.T.D. (e-mail: pdoran@uic.edu).

DNA reveals high dispersal synchronizing the population dynamics of Canada lynx

Michael K. Schwartz^{*†}, L. Scott Mills^{*}, Kevin S. McKelvey[‡], Leonard F. Ruggiero[‡] & Fred W. Allendorf[‡]

^{*} Wildlife Biology Program, School of Forestry, University of Montana, Missoula, Montana 59812, USA

[†] USDA/USFS Rocky Mountain Research Station, 800 E. Beckwith, Missoula, Montana 59801, USA

[‡] Division of Biological Sciences, University of Montana, Missoula, Montana 59812, USA

Population dynamics of Canada lynx (*Lynx canadensis*) have been of interest to ecologists for nearly sixty years^{1–4}. Two competing hypotheses concerning lynx population dynamics and large-scale spatial synchrony are currently debated. The first suggests that dispersal is substantial among lynx populations⁵, and the second proposes that lynx at the periphery of their range exist in small, isolated patches that maintain cycle synchrony via correlation with extrinsic environmental factors². Resolving the nature of lynx population dynamics and dispersal is important both to ecological theory and to the conservation of threatened lynx populations: the lack of knowledge about connectivity between populations at the southern periphery of the lynx's geographic range delayed their legal listing in the United States⁶. We test these competing hypotheses using microsatellite DNA markers and lynx samples from 17 collection sites in the core and periphery of the lynx's geographic range. Here we show high gene flow despite separation by distances greater than 3,100 km, supporting the dispersal hypothesis. We therefore suggest that management actions in the contiguous United States should focus on maintaining connectivity with the core of the lynx's geographic range.

Trapping records show that twentieth-century lynx population dynamics in parts of North America exhibit patterns of lagged synchrony, with irruptions occurring in the centre of the continent 2–3 years before they occur at the periphery of the range⁵. One potential mechanism for this phenomenon is a travelling wave of lynx emanating from the centre of the continent, which synchronizes the lynx populations⁴. If this 'dispersal hypothesis' is correct, then populations at the periphery of their range should be highly influenced by lynx periodically diffusing outwards from the core.

Alternatively, lynx populations at the periphery of their geographic range may be self-sustaining, and largely isolated⁷. Under this 'peripheral isolation hypothesis' there would be few dispersers and the number of dispersers would decline exponentially with distance, leading to lower connectivity at the edge of the geographic range³. To support the peripheral isolation hypothesis there are weak correlations between lynx irruptions in Alberta and British Columbia and lynx abundance indices in some southern, peripheral populations, as well as the patchy nature of lynx habitat in southern Canada, Montana, and Washington^{5,7}. If populations are relatively isolated, synchrony between peripheral lynx populations could be generated by exogenous density-independent events such as weather² (that is, the Moran effect⁸).

We used nine microsatellite loci to estimate gene flow among lynx populations (see Methods). We analysed lynx samples from 17 collection sites in the periphery and core of the lynx's geographic range (Fig. 1). If the dispersal hypothesis is correct, gene flow should be high among populations, including central Canadian populations and the populations on the periphery of the geographic range. Alternatively, if the peripheral isolation hypothesis is correct, then gene flow should be low between peripheral lynx populations and

central Canadian and Alaskan populations, and negligible among populations that are far apart.

The global F_{st} , a measure of population subdivision, was 0.033 (standard error of the mean, s.e.m. ± 0.002). This degree of subdivision is expected if there are on average approximately six dispersers ('migrants' in the genetic sense) entering each population each generation, assuming an island model of migration^{9,10}. Furthermore, substantial gene flow was apparent among all populations. The Kenai Peninsula population was genetically most divergent from other populations with a mean pairwise F_{st} of 0.051 (s.e.m. ± 0.003). However, this amount of subdivision still represents approximately four dispersers entering each population per generation, and so the Kenai Peninsula population is probably not biologically different. Despite sampling lynx populations more than 3,100 km apart, we found no evidence for decreased gene flow with increasing geographical distance across western North America (Fig. 2; Mantel's test, $g = 0.117$, $P = 0.42$).

Small F_{st} values can be indicative of high current gene flow between populations or can be caused by populations sharing recent common ancestry⁹. We attribute our results to high current gene flow because many peripheral populations in our study have had small population sizes for long periods. Lynx are known to have low population densities^{5,7}, especially at cyclic lows that would reduce effective population size (N_e). On the Kenai Peninsula, our estimate of N_e was less than 30 (see Methods). For ideal isolated populations with $N_e = 30$, substantial values of F_{st} would accumulate in only a few generations (t). For example, F_{st} is expected to be greater than the global F_{st} of 0.033 in just two generations for populations with an $N_e = 30$, and in four generations for populations with $N_e = 50$ ($F_{st} = 1 - (1 - 1/(2N_e))^t$ ⁹⁻¹¹).

Our F_{st} results are corroborated by assignment test results (see Methods)¹². Only 40.8% of lynx assigned to the population from which they were captured. Low assignment rates may indicate either

high gene flow or low power to assign because of too few markers or too little genetic variation per marker. However, other studies with less overall genetic variation and equal numbers of microsatellites have produced much higher assignment rates¹³, so we attribute our low assignment rates to high gene flow.

Radiotelemetry data have shown that lynx regularly travel distances greater than 100 km, and can travel distances up to 1,100 km (refs 14, 15). However, it is unknown whether these movements led to gene flow. Our genetic data suggest that long distance movements are probably common and result in very high levels of gene flow, among the highest yet found for any carnivore. Wolves and coyotes show high levels of gene flow¹⁶, yet wolves still follow an isolation by distance model¹⁷. North American brown bears also display high gene flow, but have F_{st} values much higher (implying gene flow levels much lower) than are reported here for lynx¹⁸.

Our results for lynx strongly support the dispersal hypothesis rather than the peripheral isolation hypothesis. Peripheral populations in the south, north and west appear to readily exchange dispersers with the core populations. Even the peninsular Kenai population shows high gene flow.

Gene flow has implications for synchrony in lynx cycles across large landscapes. Stenseth *et al.*² used lynx fur trade records from Canada's Hudson Bay Company along with time-series data from Statistics Canada to show that density-independent factors (that is, weather) synchronize isolated lynx populations with similar density-dependent structures. Specifically, their models based on climatic regions (Pacific maritime, Continental, and Atlantic maritime) had more support than models subdividing lynx populations based on ecological groupings (western, northern, southern and eastern), provincial boundaries, or Hudson Bay Company administrative regions. Stenseth *et al.*² concluded that region-specific variation in climate, probably produced by the North Atlantic Oscillation, coupled with similar density-dependent structures in lynx populations, caused lynx cycling synchrony within climatic regions. A different model³ explained large-scale spatial synchrony by assuming that dispersal between patches declined exponentially with distance.

We suggest that immediately after the peak of the lynx cycle in the centre of their range, large numbers of lynx disperse long distances creating a wave of immigrants that drive cycle-like synchrony in the western lynx populations. This suggestion is supported by both trapping records⁵ and our gene flow results. A dispersal hypothesis is also a more parsimonious explanation of lynx cycle synchrony, negating the reliance on large-scale density-independent events coupled with similar density-dependent population structures¹⁹⁻²². Dispersal may also be significant in synchronizing population cycles in other species. For example, initial research on collared lemmings showed that synchrony occurred only in populations separated by as much as 6 km (ref. 23). Because this distance was greater than the

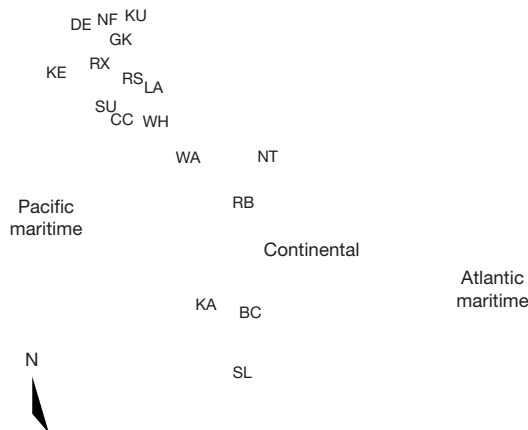


Figure 1 The geographic distribution of lynx (in grey). Each triangle represents a location of a 'peripheral' lynx population and each circle represents a 'core' lynx population. The population abbreviations and sample sizes are as follows: Seeley Lake, Montana (SL, 32); Kuyuktuvuk Creek, Alaska (KU, 7); Kenai Peninsula, Alaska (KE, 115); Ladue River, Yukon/Alaska (LA, 10); Cooper Center, Alaska (CC, 19); North of Fairbanks, Alaska (NF, 19); Fort Providence, Northwest Territories (NT, 84); Riverside, Alaska (RS, 43); West of Denali, Alaska (DE, 16); Rainbow Lake, Alberta (RB, 18); Watson Lake, Yukon (WA, 27); Whitehorse, Yukon (WH, 52); Kootnay-Banff, British Columbia (BC, 20); Gold King Creek, Alaska (GK, 32); North of Kamloops, British Columbia (BC, 20); Paxson, Alaska (PX, 45); and Susitna Lake, Alaska (SU, 35). The regions Atlantic, Continental and Pacific represent the climatic regions of ref. 2.

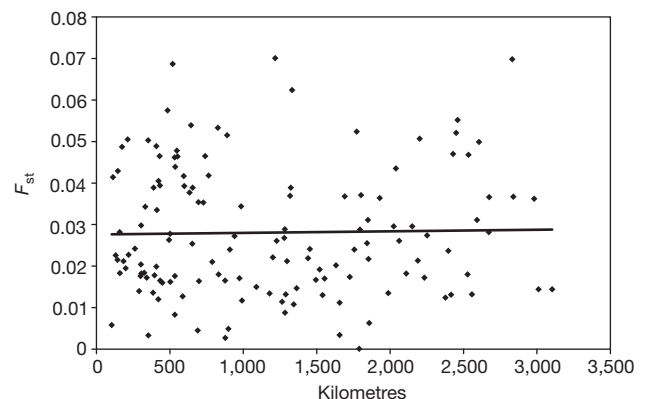


Figure 2 Regression of F_{st} on geographic distance between all pairs of the 17 populations.

maximum observed lemming dispersal distance (3 km) (ref. 24), dispersal was discounted as a synchronizing mechanism. However, recent genetic data revealed that collared lemming disperse distances up to 20 km (ref. 24), suggesting that dispersal may indeed synchronize these populations.

Our results have important implications for lynx conservation. Our data imply that persistence in the contiguous United States depends upon dispersal from larger populations; therefore joint international efforts should be initiated to ensure that connectivity between northern and southern populations is sustained. □

Method

Genetic analysis

We genotyped 599 lynx samples from 17 populations using nine microsatellite DNA markers developed from domestic cats^{25,26}. DNA extraction methods, microsatellite DNA amplification conditions, and Hardy–Weinberg (HW) proportions and gametic disequilibrium analyses can be found in ref. 27. Average heterozygosity across all populations and loci was 0.66 (s.e.m. = 0.074). Several populations had one locus out of HW proportions ($P < 0.05$); however, there was no consistency as to which locus. The only population with more than one locus out of HW proportions was the Kenai population that had three of nine loci deviating from HW proportions.

Effective population size

We estimated N_e of the Kenai lynx population using the temporal change in allele frequency method²⁸. Our samples were collected 10 years apart, a period representing between two and three lynx generations. Assuming the samples were separated by two generations produced an N_e estimate of 22.1 (s.e.m. = 11.5–49.1); assuming the samples were separated by three generations resulted in an N_e estimate of 28.8 (s.e.m. = 17.6–62.0).

Assignment tests

An assignment test classified an individual to a population where it most probably was born, on the basis of the expected frequency of an individual's genotype in each population¹². We used the partially bayesian exclusion test of ref. 12, which has been shown to be effective over a wide range of F_{st} values and is robust to slight deviations from Hardy–Weinberg proportions^{12,13}. We used the 'leave one out' method when conducting this analysis, which means that each individual was removed from the data set, the allele frequencies were recalculated, and then the individual was assigned to the population.

Received 5 September; accepted 18 November.

- Elton, C. S. & Nicholson, M. The ten-year cycle in numbers of the lynx in Canada. *J. Anim. Ecol.* **11**, 215–244 (1942).
- Stenseth, N. C. *et al.* Common dynamic structure of Canada lynx populations within three climatic regions. *Science* **285**, 1071–1073 (1999).
- Ranta, E., Kaitala, V. & Lunnberg, P. The spatial dimension in population fluctuations. *Science* **278**, 1621–1623 (1997).
- Blasius, B., Huppert, A. & Stone, L. Complex dynamics and phase synchronization in spatially extended ecological systems. *Nature* **399**, 354–359 (1999).
- McKelvey, K. S., Aubry, K. B. & Ortega, Y. K. in *Ecology and Conservation of Lynx in the United States* (eds Ruggiero, L. F. *et al.*) 207–264 (Univ. Press of Colorado, Denver, 2000).
- Federal Register. Endangered and threatened wildlife and plants; determination of threatened status for the contiguous U.S. distinct population segment of the Canada lynx and related rule; final rule. *Federal Register* **65**, 16051–16086 (2000).
- Ruggiero, L. F. *et al.* in *Ecology and Conservation of Lynx in the United States* (eds Ruggiero, L. F. *et al.*) 443–454 (Univ. Press of Colorado, Denver, 2000).
- Moran, P. A. P. The statistical analysis of the Canadian lynx cycle. II. Synchronization and meteorology. *Aust. J. Zool.* **1**, 291–298 (1953).
- Wright, S. *Evolution and the Genetics of Populations* Vol. 2 *The Theory of Gene Frequencies* (Univ. Chicago Press, Chicago, 1969).
- Whitlock, M. & McCauley, D. Indirect measures of gene flow and migration: F_{st} not equal to $1/(4Nm+1)$. *Heredity* **82**, 117–125 (1999).
- Waples, R. S. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *J. Hered.* **89**, 438–450 (1998).
- Cornuet, J., Piry, S., Luikart, G., Estoup, A. & Solignac, M. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**, 1989–2000 (1999).
- Manuel, S., Bertier, P. & Luikart, G. Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. *Cons. Biol.* (in the press).
- Slough, B. G. & Mowat, G. Lynx populations dynamics in an untrapped refugium. *J. Wildl. Mgmt* **60**, 946–961 (1996).
- Mowat, G., Poole, K. G. & O'Donoghue, M. in *Ecology and Conservation of Lynx in the United States* (eds Ruggiero, L. F. *et al.*) 265–306 (Univ. Press of Colorado, Denver, 2000).
- Roy, M. S., Geffin, E., Smith, D., Ostrander, E. A. & Wayne, R. K. Patterns of differentiation and hybridization in North American wolf-like canids revealed by analysis of microsatellite loci. *Mol. Biol. Evol.* **11**, 553–570 (1994).
- Forbes, S. H. & Boyd, D. Genetic structure and migration in native and re-introduced Rocky Mountain wolf populations. *Cons. Biol.* **11**, 1226–1234 (1997).
- Paetkau, D., Shields, G. F. & Strobeck, C. Gene flow between insular, coastal and interior populations of brown bears in Alaska. *Mol. Ecol.* **7**, 1283–1292 (1998).
- Hudson, P. J. & Cattadori, I. M. The Moran effect: a cause of synchrony. *Trends Ecol. Evol.* **14**, 1–2 (1999).

- Grenfell, B. T. *et al.* Noise and determinism in synchronized sheep dynamics. *Nature* **394**, 674–677 (1998).
- Cattadori, I. M., Merler, S. & Hudson, P. J. Searching for mechanisms of synchrony in spatially structured gamebird populations. *J. Anim. Ecol.* **69**, 620–638 (2000).
- Greenman, J. V. & Benton, T. G. The impact of stochasticity on the behaviour of nonlinear population models: synchrony and the Moran effect. *Oikos* **93**, 343–351 (2001).
- Predavec, M., Krebs, C. J., Danell, K. & Hyndman, R. Cycles and synchrony in the collared lemming (*Dicrostonyx groenlandicus*) in Arctic North America. *Oecologia* **126**, 216–224 (2001).
- Ehrlich, D. *et al.* Spatial structuring of lemming populations (*Dicrostonyx groenlandicus*) fluctuating in density. *Mol. Ecol.* **10**, 481–496 (2001).
- Menotti-Raymond, M. A. & O'Brien, S. J. Evolutionary conservation of ten microsatellite loci in four species of Felidae. *J. Hered.* **86**, 319–322 (1995).
- Menotti-Raymond, M. A. *et al.* A genetic linkage map of microsatellites in the domestic cat (*Felis catus*). *Genomics* **57**, 9–23 (1999).
- Schwartz, M. K., Mills, L. S., Allendorf, F. W. & Ruggiero, L. F. Landscape location affects genetic variation of Canada lynx (*Lynx canadensis*). *Mol. Ecol.* (submitted).
- Berthier, P., Beaumont, M., Cornuet, J.-M. & Luikart, G. Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics* (in the press).

Acknowledgements

We thank the following people and institutions for providing samples: C. Apps, T. Bailey, H. Golden, G. Jarrell, J. Cook, M. Hebblewhite, Montana Fish Wildlife and Parks, J. Kolbe, R. Mulders, B. Naney, R. Oakleaf, L. Roy, B. Scotton, T. Shurry, H. Slama and J. Squires. We thank H. Draheim, B. Adams, B. Theroux, S. Forbes, P. Spruell and K. Pilgrim for laboratory support and advice. This project was funded by the USDA/USFS (grant to L.S.M. and M.K.S.), and NSF (grant to L.S.M.); M.K.S. was additionally funded by a McIntire-Stennis grant, the USFS Rocky Mountain Research Station, and the NSF Training-WEB. We thank R. Bick, S. Forbes, G. Luikart, D. Pletscher, M. Poss, D. Tallmon and E. Winer for comments on earlier versions of this manuscript. All pertinent local, national and international permits required for this project are on file at the University of Montana.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.K.S. (e-mail: mks@selway.umn.edu).

Identification of diploid endosperm in an early angiosperm lineage

Joseph H. Williams*† & William E. Friedman*†

* Department of Environmental, Population and Organismic Biology, University of Colorado, Boulder, Colorado 80309, USA

† These authors contributed equally to the work

In flowering plants, the developmental and genetic basis for the establishment of an embryo-nourishing tissue differs from all other lineages of seed plants. Among extant nonflowering seed plants (conifers, cycads, *Ginkgo*, Gnetales), a maternally derived haploid tissue (female gametophyte) is responsible for the acquisition of nutrients from the maternal diploid plant, and the ultimate provisioning of the embryo. In flowering plants, a second fertilization event, contemporaneous with the fusion of sperm and egg to yield a zygote, initiates a genetically biparental and typically triploid embryo-nourishing tissue called endosperm. For over a century, triploid biparental endosperm has been viewed as the ancestral condition in extant flowering plants^{1–3}. Here we report diploid biparental endosperm in *Nuphar polysepalum*, a basal angiosperm. We show that diploid endosperms are common among early angiosperm lineages and may represent the ancestral condition among flowering plants. If diploid endosperm is plesiomorphic, the triploid endosperms of the vast majority of flowering plants must have evolved from a diploid condition through the developmental modification of the unique fertilization process that initiates endosperm.

In 1999, a series of phylogenetic analyses^{4–6} identified a set of

flowering plant clades that diverged before the origin of monocots or eudicots. This basal grade of angiosperm taxa includes monotypic *Amborella trichopoda*, monophyletic Nymphaeales (water lilies), and a clade that includes Illiciaceae, Schisandraceae, Trimeniaceae and Austrobaileyaceae (the ITA clade). These, and more recent phylogenetic analyses consistently indicate that *Amborella* or *Amborella* plus Nymphaeales is the sister group to all other angiosperms^{4–8}.

This identification of the most basal angiosperm lineages has begun to alter long-held suppositions about the biological features of the earliest flowering plants, such as the primitive (plesiomorphic) morphology of the flower⁹ and the anatomy of water-conducting tissues¹⁰. We know very little about two of the most characteristic features of angiosperms, double fertilization and endosperm, in early flowering plant lineages. There are no published data that conclusively demonstrate a second fertilization event or report the genetic constitution of endosperm in these basal taxa^{11,12}. Despite the importance of endosperm in the life cycles of flowering plants (and of humans, as two thirds of human caloric intake worldwide derives from endosperm¹³), the origin and early evolution of this unusual tissue remain unclear.

We performed a developmental analysis of the fertilization process in *Nuphar*, an early divergent lineage within the Nymphaeaceae (one of the two families that comprise the Nymphaeales). In *Nuphar*, the mature egg-producing structure (female gametophyte) contains four cells: an egg cell, a central cell and two sterile cells (synergids) (Fig. 1). Each of these cells, including the central cell, contains one nucleus. Most angiosperms have female gametophytes in which eight nuclei are partitioned into seven cells: a uninucleate egg cell, a binucleate central cell, and five additional uninucleate cells (two synergids and three antipodals). Because the initiation of endosperm typically involves the fusion of a sperm with the nuclear contents of the central cell of the female gametophyte, the pattern of

development of the female gametophyte directly determines the ultimate genetic constitution and ploidy level of the endosperm. Among angiosperms, the central cell may contain up to 14 (typically haploid) nuclei, but two nuclei is by far the most common condition^{14,15}.

Four-celled female gametophytes in *Nuphar*¹⁶ and other recently recognized basal angiosperms have been reported previously¹⁵. However, the data from many of these studies (typically in the form of drawings) have been questioned^{14,17}. This is partly because the female gametophytes of many flowering plants with seven cells and eight nuclei may initiate programmed cell death of the three sterile antipodal cells and the two haploid nuclei of the central cell may fuse before fertilization¹⁵. Thus an originally seven-celled female gametophyte may become structurally indistinguishable, at maturity, from a 'true' four-celled female gametophyte^{2,14} (Fig. 2). However, the central cell of a 'cryptic' seven-celled female gametophyte will be diploid, and if fertilized, will yield a triploid biparental endosperm, whereas a 'true' four-celled female gametophyte will contain a haploid central cell that will, if fertilized, yield a diploid biparental endosperm (Fig. 2).

We used microspectrofluorometric techniques in combination with DAPI, a DNA-binding fluorochrome, to determine absolute ploidy levels of the egg cell nucleus, central cell nucleus, zygote nucleus and primary endosperm nucleus (the nucleus of endosperm at the one-celled stage of development) in *Nuphar*. In order to calibrate relative fluorescence per C unit of DNA, we measured diploid zygote nuclei during mitosis, which contain, by definition, the 4C content of DNA. Fluorescence of DAPI-stained zygote nuclei in mitosis averaged 40.44 ± 2.30 (mean $\pm$ standard deviation; $n = 5$) relative fluorescence units (RFU). These data allow for the calibration of relative fluorescence per C unit of DNA and indicate that in *Nuphar* a 4C quantity of DNA corresponds quite closely to 40 RFU; a 2C quantity of DNA corresponds to 20 RFU; and that 10 RFU reflect a 1C quantity of DNA.

We observed fertilization events in ovules of *N. polysepalum* collected at 13 and 22 h after pollination. In 12 unfertilized ovules (in which no pollen tubes were present 22 h after pollination), the mean fluorescence of egg and central cell nuclei were 10.19 ± 0.93 RFU and 10.31 ± 1.36 RFU, respectively. These values are statistically indistinguishable ($P = 0.82$, paired t -test) and correspond almost exactly to the predicted 1C level of DNA fluorescence based on zygote mitotic figures (10.11 RFU). Thus, the egg nucleus and central cell nucleus are haploid.

In *Nuphar*, the two sperm that will participate in double fertilization (to form an embryo and an endosperm) are first deposited into the female gametophyte by a pollen tube. One sperm nucleus fuses with the egg nucleus. The second sperm nucleus approaches and fuses with the central cell nucleus to form the primary endosperm nucleus (Fig. 3a–d). Immediately after fusion of gametic nuclei (based on

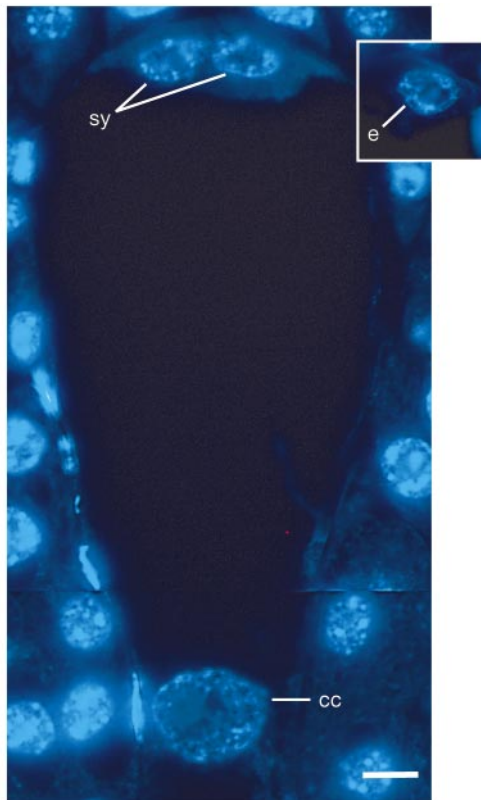


Figure 1 Four-celled/four-nucleate *Nuphar* female gametophyte. Two synergids (sy) and an egg (e) are located at the micropylar end of the female gametophyte. The egg cell (inset) is in a separate section directly behind the synergids. The central cell is defined by the dark region below the egg and synergids and contains a single nucleus (cc). Composite photo in main panel from two serial sections. Scale bar, 10 μ m.

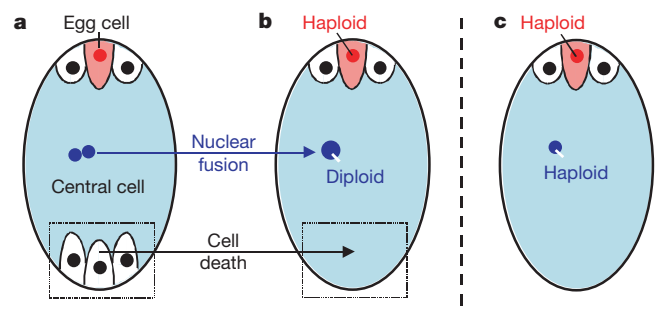


Figure 2 Convergence in form of angiosperm female gametophytes. **a, b**, A seven-celled, eight-nucleate female gametophyte (**a**) undergoes degeneration of three sterile cells (antipodals) and fusion of the two nuclei of the central cell to yield a four-celled female gametophyte with diploid central cell (**b**). **c**, A 'true' four-celled female gametophyte with a single haploid nucleus in the central cell, as in *Nuphar*.

visual evidence of the presence of sperm chromatin in contact with or just within the nuclear envelope of the egg or central cell nucleus; Fig. 3a–c), the zygote nucleus and primary endosperm nucleus possess equivalent relative DNA contents (mean zygote nucleus, 19.73 ± 2.55 RFU; mean primary endosperm nucleus, 20.33 ± 1.73 RFU; $P = 0.60$; paired t -test; $n = 7$ pairs). These means are almost exactly twice the means for unfertilized egg and central cell nuclei and correspond to the predicted 2C level of DNA fluorescence (20.22 RFU). These data represent the first definitive evidence, to our knowledge, of a second fertilization event in a basal angiosperm.

Finally, we collected relative DNA fluorescence data from primary endosperm nuclei in mitosis. The mean relative fluorescence of mitotic primary endosperm nuclei, 40.94 ± 2.07 RFU ($n = 4$), is almost identical to the mean fluorescence of zygote nuclei in mitosis (40.44 ± 2.30 RFU), which contain the 4C quantity of DNA. Thus, both the zygote and primary endosperm nucleus progress through the synthesis phase of the cell cycle between fertilization and the first mitotic division (see Supplementary Information). We also measured the relative fluorescence of mitotic figures from two-celled endosperms (mean = 42.39 ± 2.74 RFU; $n = 7$). The mean fluorescence values for mitotic zygote nuclei ($n = 5$) and mitotic early endosperm nuclei ($n = 11$) are not significantly different from each other ($P = 0.33$; t -test).

Taken together, our data clearly demonstrate three essential facts about the reproductive process in *N. polysepalum*: (1) the mature female gametophyte is four-celled and possesses a haploid uninucleate central cell (and egg cell) in G1 of the cell cycle until fertilization; (2) the newly formed zygote nucleus and primary-endosperm nucleus both occupy the G1 phase of the cell cycle; and (3) this basal angiosperm initiates a diploid biparental endosperm.

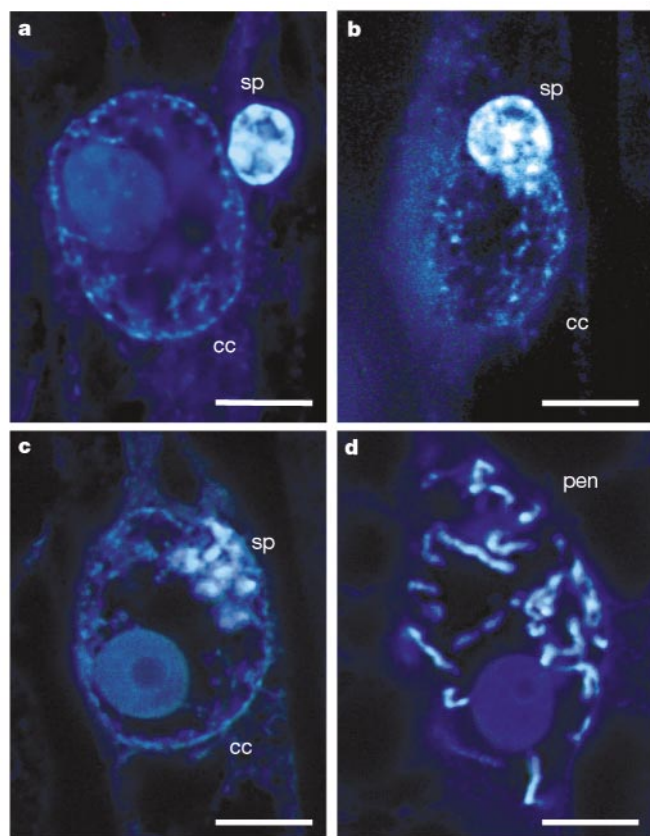


Figure 3 The second fertilization event in *Nuphar*. **a–c**, Sections from four separate ovules at 13 and 22 h after pollination. A sperm nucleus (sp) approaches (**a**), begins to fuse with (**b**), and is engulfed by (**c**) the haploid central cell nucleus (cc) to complete the fertilization event that initiates a diploid endosperm. **d**, A primary endosperm nucleus (pen, 22 h after pollination) in early prophase. Scale bar, 10 μ m.

In four other currently recognized basal angiosperm genera (*Nymphaea*, *Cabomba*, *Schisandra*, *Illicium*), previous studies^{18–25} provide indirect developmental evidence consistent with the presence of four-celled female gametophytes. Our developmental studies confirm these reports of four-celled female gametophytes for *Cabomba*, *Schisandra* and *Illicium* (data not shown). Thus, we suggest conservatively that four basal angiosperm genera, in addition to *Nuphar*, are likely to produce diploid endosperms. In contrast, a recent study of *Amborella* clearly shows a seven-celled female gametophyte with a binucleate central cell²⁶ that should yield a triploid endosperm, if a second fertilization event occurs in this taxon (the fertilization process in *Amborella* is unstudied).

We mapped this comparative information onto current phylogenies of basal angiosperms (*Amborella* or *Amborella* plus *Nymphaeales* as sister to all other angiosperms). The results show that it

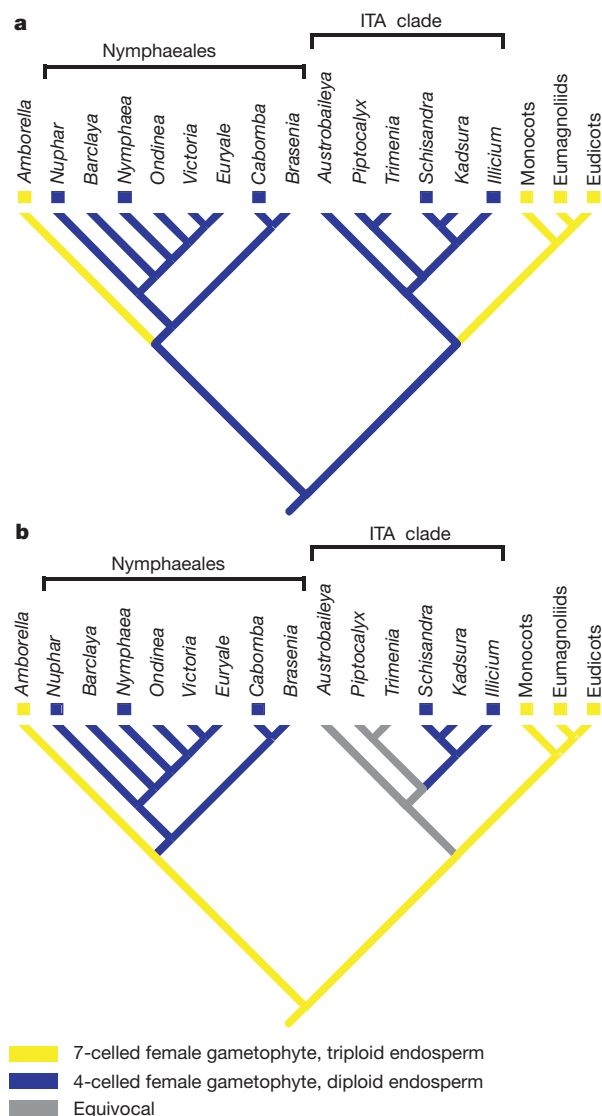


Figure 4 Evolution of female gametophyte structure and ploidy level of endosperm in basal angiosperms. Phylogeny based on recent published analyses (see text). Clades for which plesiomorphic character states are known are indicated with a colour-coded box under the taxon name. Endosperm ploidy is known for *Nuphar*, many monocots and eudicots, and is inferred (on the basis of female gametophyte structure) for *Amborella*, *Nymphaea*, *Cabomba*, *Schisandra*, *Illicium* and eumagnoliids. The plesiomorphic condition for angiosperms is unresolved and could be either diploid or triploid. Under the constraints of this phylogenetic hypothesis, either triploid endosperm evolved twice from a diploid condition during the early evolution of angiosperms (**a**) or diploid endosperm evolved twice from a triploid condition (**b**).

is equivocal whether diploid or triploid endosperm is ancestral among extant angiosperms (Fig. 4). For example, it is equally parsimonious to derive a diploid endosperm twice (once in Nymphaeales and once in the ITA clade) from an ancestral triploid endosperm or to derive a triploid endosperm twice (once in *Amborella* and once in the common ancestor of monocots, eumagnoliids and eudicots) from an ancestral diploid endosperm (Fig. 4). The century-old paradigm that extant lineages of early angiosperms have triploid endosperm¹¹ can no longer be favoured. In addition, the long-held view that an eight-nucleate, seven-celled (*Polygonum*-type) female gametophyte is primitive among angiosperms must also be questioned.

Evolutionary biologists have speculated that a diploid biparental embryo-nourishing tissue was a transitional stage between the strictly maternal haploid embryo-nourishing female gametophyte of nonflowering seed plants and the triploid biparental endosperm of most extant flowering plants^{11,27}. This hypothesis is supported by theoretical analyses of the origin of the angiosperm reproductive syndrome based on inclusive fitness constructs, parent-offspring conflict and intersexual genetic conflict^{11,28}.

The specific hypothesis that a diploid embryo was the evolutionary antecedent and homologue of endosperm was bolstered in 1992 by the discovery of a regular second fertilization event that yields a supernumerary embryo in members of Gnetales¹, a group of nonflowering seed plants. However, recent molecular phylogenetic evidence indicates that Gnetales is much more distantly related to flowering plants than was previously hypothesized. Thus, double-fertilization processes in these two lineages may well be homoplasious¹¹.

The presence of diploid endosperm in basal angiosperms suggests an evolutionary historical pathway from the haploid genetically maternal embryo-nourishing tissues of nonflowering seed plants to the triploid genetically biparental endosperm of most angiosperms. The first step involved the establishment of a new fertilization process to produce a genetically biparental diploid tissue that acquired embryo-nourishing capacity. The ancestral embryo-nourishing function of the haploid female gametophyte (characteristic of all nonflowering seed plants) was superseded by this sexually produced diploid endosperm and the female gametophyte was reduced to the four-celled form characteristic of basal angiosperms. Finally, the triploid endosperms of the vast majority of flowering plants evolved through the addition of a second haploid female nucleus to the fertilization process that initiates endosperm; this transition may have occurred once or even twice during the earliest phases of angiosperm evolution.

Recent phylogenetic hypotheses suggest far greater evolutionary distances between flowering plants and all other extant seed plants than had previously been imagined^{29,30}. Thus, the task of determining the homologies and evolutionary histories of defining angiosperm characters, in essence solving Darwin's "abominable mystery", appears as daunting as ever. If diploid endosperm represents the ancestral condition for flowering plants, a key intermediate condition in the early evolutionary history of angiosperms has been revealed. This finding would be congruent with earlier analyses that suggested an origin of endosperm from a diploid condition and specifically posited an embryo as the homologue of endosperm^{1,28}. The presence of diploid endosperm in an early angiosperm lineage brings us one step closer to bridging the substantial gap between flowering plants and their seed plant ancestors. □

Methods

Plant collection

Reproductive material of *N. polysepalum* was collected in June and July of 2000 and 2001 at Red Rocks Lake (Brainard Lake Recreation area), Colorado, USA. Supplementary pollinations were performed on a subset of flowers in the field and after flowers had been transported to the laboratory. Ovules were dissected from flowers and chemically fixed between one and 53 h after pollination.

Histology

Flowers were chemically fixed for 24 h in either 3:1 95% ethanol:acetic acid or 4% acrolein in a solution of 50 mM PIPES buffer (also 5 mM EGTA and 1 mM MgSO₄; pH 6.8). Specimens were dehydrated through an ethanol series, infiltrated and embedded with glycol methacrylate, and serially sectioned into 5-μm-thick ribbons. Flowers fixed in acrolein were stained in 0.1% toluidine blue. Flowers fixed in the 3:1 solution were stained with a solution of 0.25 μg ml⁻¹ of 4',6-diamidino-2-phenylindole (DAPI) and 0.1 mg ml⁻¹ p-phenylenediamine in 0.05 M Tris (pH 7.2). Digital micrographs were taken on a Zeiss Axiophot microscope equipped with a Zeiss Axiocam digital camera, and images were processed with Adobe Photoshop software.

Microspectrofluorometry

Sections from flowers fixed in the 3:1 solution were stained with DAPI (see above) for 1 h at room temperature in a light-free environment. Microspectrofluorometric measurements of relative DNA levels of DAPI-stained nuclei were performed within 2 h. Measurements were made with a Zeiss MSP 20 microspectrophotometer with digital microprocessor coupled to a Zeiss Axiocam microscope equipped with epifluorescence (HBO 100-W burner). An ultraviolet filter set (model number 48702) with excitation filter (365 nm, bandpass 12 nm), dichroic mirror (FT395), and barrier filter (LP397) was used with a Zeiss Plan Neofluar 20× objective. Before each recording session, the photometer was standardized by taking a reading of fluorescence emitted from a fluorescence standard (GG 17), and this reading was taken to represent 100 relative fluorescence units (RFU). At the completion of each session, an additional reading was made of the fluorescence standard to confirm that little or no deviation in the relative fluorescence value obtained from the fluorescence standard had occurred during the period of data recording. Relative DNA content for each nucleus was determined by summing the individual fluorescence values of each of the serial sections through that nucleus. A net photometric value for each section of a nucleus was obtained by recording an initial reading of the nucleus and subtracting a background value obtained from cytoplasm proximal to the nucleus. Thus, background fluorescence from the glycol methacrylate was removed from the photometric analysis of relative DNA content.

Received 3 August; accepted 19 November 2001.

- Friedman, W. E. Organismal duplication, inclusive fitness theory, and altruism: Understanding the evolution of endosperm and the angiosperm reproductive syndrome. *Proc. Natl Acad. Sci. USA* **92**, 3913–3917 (1995).
- Palser, B. E. The bases of angiosperm phylogeny: embryology. *Ann. Missouri Bot. Gard.* **62**, 621–646 (1975).
- Stebbins, G. L. *Flowering Plants: Evolution Above the Species Level* (Harvard Univ. Press, Cambridge, Massachusetts, 1974).
- Mathews, S. & Donoghue, M. J. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* **286**, 947–950 (1999).
- Qui, Y.-L. *et al.* The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**, 404–407 (1999).
- Soltis, P. S., Soltis, D. E. & Chase, M. W. Angiosperm phylogeny inferred from multiple genes as a research tool for comparative biology. *Nature* **402**, 402–404 (1999).
- Graham, S. W. & Olmstead, R. G. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *Am. J. Bot.* **87**, 1712–1730 (2000).
- Barkman, T. J., Chenery, G., McNeal, J. R., Lyons-Weiler, J. & dePamphilis, C. W. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proc. Natl Acad. Sci. USA* **97**, 13166–13171 (2000).
- Endress, P. K. The flowers in extant basal angiosperms and inferences on ancestral flowers. *Int. J. Plant Sci.* **162**, 1111–1140 (2001).
- Field, T. S. *et al.* Structure and function of tracheary elements in *Amborella trichopoda*. *Int. J. Plant Sci.* **161**, 705–712 (2000).
- Friedman, W. E. & Floyd, S. K. The origin of flowering plants and their reproductive biology—a tale of two phylogenies. *Evolution* **55**, 217–231 (2001).
- Friedman, W. E. Comparative embryology of basal angiosperms. *Curr. Opin. Plant Biol.* **4**, 14–20 (2001).
- Cassman, K. G. Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agriculture. *Proc. Natl Acad. Sci. USA* **96**, 5952–5959 (1999).
- Maheshwari, P. *An Introduction to the Embryology of Angiosperms* (McGraw-Hill, New York, 1950).
- Johri, B. M., Ambegaokar, K. B. & Srivastava, P. S. *Comparative Embryology of Angiosperms* (Springer, Berlin, Germany, 1992).
- Winter, A. N. & Shamrov, I. I. Development of the ovule and embryo sac in *Nuphar lutea* (Nymphaeaceae). *Bot. Zhurnal* **76**, 378–390 (1991).
- Battaglia, E. The evolution of the female gametophyte of angiosperms: an interpretative key. (Embryological Questions 14). *Ann. Bot.* **47**, 7–144 (1989).
- Orban, I. & Bouharmont, J. Megagametophyte development of *Nymphaea nouchali* Burm. f. (Nymphaeaceae). *Bot. J. Linn. Soc.* **126**, 339–348 (1998).
- van Miegroet, F. & Dujardin, M. Cytologie et histologie de la reproduction chez le *Nymphaea heudelottii*. *Can. J. Bot.* **70**, 1991–1996 (1992).
- Winter, A. N. & Shamrov, I. I. Megasporogenesis and embryo sac development in representatives of the genera *Nymphaea* and *Victoria* (Nymphaeaceae). *Bot. Zhurnal* **76**, 1716–1728 (1991).
- Batygina, T. B., Shamrov, I. I. & Kolesova, G. E. Embryology of the Nymphaeales and Nelumbonales II. The development of the female embryonic structures. *Bot. Zhurnal* **67**, 1179–1195 (1982).
- Galati, B. G. Estudios embriológicos en *Cabomba australis* (Nymphaeaceae) I. La esporogénesis y las generaciones sexuales. *Boletín Soc. Argentina Bot.* **24**, 29–47 (1985).
- Swamy, B. G. L. Macrogametophytic ontogeny in *Schisandra chinensis*. *J. Indian Bot. Soc.* **43**, 391–396 (1964).
- Yoshida, O. Embryologische studien über *Schisandra chinensis* Bailey. *J. Coll. Arts Sci. Chiba Univ.* **3**, 459–462 (1962).

25. Solntseva, M. P. in *Comparative Embryology of Flowering Plants* (ed. Yakovlev, M. S.) 51–54 (Nauka, Leningrad, Russia, 1981) (in Russian).
26. Tobe, H., Jaffre, T. & Raven, P. H. Embryology of *Amborella* (Amborellaceae): descriptions and polarity of character states. *J. Plant Res.* **113**, 271–280 (2000).
27. Wilson, M. F. & Burley, N. *Mate Choice in Plants: Tactics, Mechanisms, and Consequences* (Princeton Univ. Press, Princeton, 1983).
28. Queller, D. C. in *Oxford Surveys in Evolutionary Biology* (eds Harvey, P. H. & Partridge, L.) 73–109 (Oxford Univ. Press, Oxford, 1989).
29. Chaw, S. M., Parkinson, C. L., Cheng, Y., Vincent, T. M. & Palmer, J. D. Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proc. Natl Acad. Sci. USA* **97**, 4086–4091 (2000).
30. Bowe, L. M., Coat, G. & dePamphilis, C. W. Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc. Natl Acad. Sci. USA* **97**, 4092–4097 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank P. K. Diggle, L. Hufford and R. H. Robichaux for critical comments on the manuscript and W. Gallup and M. Dozier for assistance with histology. This work was supported by grants from the National Science Foundation to W.E.F.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.E.F. (e-mail: ned@colorado.edu).

Mechanism for the learning deficits in a mouse model of neurofibromatosis type 1

Rui M. Costa*, Nikolai B. Federov†, Jeff H. Kogan†, Geoffrey G. Murphy*, Joel Stern*, Masuo Ohno*, Raju Kucherlapati‡, Tyler Jacks§ & Alcino J. Silva*

* Departments of Neurobiology, Psychiatry and Psychology, BRL, University of California at Los Angeles, Los Angeles, California 90095-1761, USA

‡ Department of Molecular Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York, New York 10461, USA

§ Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Neurofibromatosis type I (NF1) is one of the most common single-gene disorders that causes learning deficits in humans¹. Mice carrying a heterozygous null mutation of the *Nf1* gene (*Nf1*^{+/-}) show important features of the learning deficits associated with NF1 (ref. 2). Although neurofibromin has several known properties and functions, including Ras GTPase-activating protein activity^{3,4}, adenylyl cyclase modulation^{5,6} and microtubule binding⁷, it is unclear which of these are essential for learning in mice and humans. Here we show that the learning deficits of *Nf1*^{+/-} mice can be rescued by genetic and pharmacological manipulations that decrease Ras function. We also show that the *Nf1*^{+/-} mice have increased GABA (γ-amino butyric acid)-mediated inhibition and specific deficits in long-term potentiation, both of which can be reversed by decreasing Ras function. Our results indicate that the learning deficits associated with NF1 may be caused by excessive Ras activity, which leads to impairments in long-term potentiation caused by increased GABA-mediated inhibition. Our findings have implications for the development of treatments for learning deficits associated with NF1.

Visual-spatial problems are among the most common cognitive deficits detected in individuals affected with NF1 (refs 1, 8). Studies have shown that *Nf1*^{+/-} mice have abnormal spatial learning² when tested in the hidden version of the water maze—a task that is sensitive to hippocampal lesions⁹. Studies suggest that an upregulation of Ras activity may account for the learning deficits in NF1 both in mice¹⁰ and humans¹¹. To test this hypothesis, we crossed the *Nf1*^{+/-} mice (C57B6/N background) with mice heterozygous for a null mutation in the *K-ras* gene¹² (*K-ras*^{+/-}, 129T2/SvEms), and tested the F₁ descendants (hybrid isogenic background) in the hidden version of the water maze task. Because *K-ras*^{+/-} mice, like *Nf1*^{+/-} (ref. 13), die *in utero*¹², we used *K-ras*^{+/-} mice, which are viable and show no apparent developmental defects.

Mice were trained with two trials per day. In all the following experiments, no differences between genotypes and/or treatments were observed in acquisition (Fig. 1a, d, g), floating, thigmotaxic behaviour or swimming speed (data not shown). Spatial learning was assessed in a probe trial¹⁴ (training day 7), in which the platform was removed from the pool. The time spent searching in the training quadrant was different among the different genotypes (analysis of variance, ANOVA, $F_{3,61} = 7.27$, $P < 0.05$). Wild-type mice spent significantly more time searching in the training quadrant than did *Nf1*^{+/-} mice (Fisher's protected least significant difference, PLSD, $P < 0.05$; Fig. 1b), confirming that the *Nf1*^{+/-} mice have impaired spatial learning. *K-ras*^{+/-} mice were also impaired ($P < 0.05$; Fig. 1b); however, mice carrying heterozygous mutations in both the *Nf1* and the *K-ras* genes (*Nf1*^{+/-}/*K-ras*^{+/-}) spent as much time as wild-type mice in the training quadrant ($P > 0.05$; Fig. 1b), and significantly more time than *Nf1*^{+/-} and *K-ras*^{+/-} mice ($P < 0.05$). This indicates that the learning deficits in *Nf1*^{+/-} mice can be rescued by the *K-ras*^{+/-} mutation. Furthermore, both wild-type (paired *t*-test, $t_{23} = 5.935$, $P < 0.05$) and *Nf1*^{+/-}/*K-ras*^{+/-} mice searched selectively for the missing platform; that is, they searched closer to the exact platform position than to the opposite position in the pool¹⁵ ($t_{10} = 4.09$, $P < 0.05$; Fig. 1c), whereas *Nf1*^{+/-} and *K-ras*^{+/-} mice did not (*Nf1*^{+/-} $t_{14} = 1.83$; *K-ras*^{+/-} $t_{14} = 0.149$; $P > 0.05$).

Mutations that decrease Ras function do rescue the deficits of the *Nf1*^{+/-} mice because a null heterozygous mutation in *N-ras* (*N-ras*^{+/-}), another mammalian *ras* gene with overlapping function and similar expression to that of *K-ras*¹², also reverses the spatial learning deficits of the *Nf1*^{+/-} mice (Fig. 1e, f; see Methods). During the probe trial, there was a significant effect of genotype on searching strategy ($F_{3,32} = 2.83$, $P < 0.05$; Fig. 1e). *Nf1*^{+/-}/*N-ras*^{+/-} mice were indistinguishable from wild type (Fisher's PLSD, $P > 0.05$), and searched significantly more in the training quadrant than did *Nf1*^{+/-} mice ($P < 0.05$). Also, analysis of proximity scores showed that both wild-type and *Nf1*^{+/-}/*N-ras*^{+/-} mice searched selectively (wild type, $t_9 = 3.621$; *Nf1*^{+/-}/*N-ras*^{+/-}, $t_8 = 5.84$; $P < 0.05$; Fig. 1f), whereas *Nf1*^{+/-} mice did not ($t_{14} = 1.83$, $P > 0.05$). The *N-ras*^{+/-} mutation by itself did not produce any detectable phenotype ($P > 0.05$; Fig. 1e), consistent with previous studies indicating that Ras signalling is less impaired in *N-ras* than in *K-ras* mutants¹².

The *ras* mutations that reversed the *Nf1*^{+/-} spatial learning deficits should result in a decrease in Ras function throughout the life of the mice. To determine whether decreases in Ras function specifically during training could reverse the learning deficits of the *Nf1*^{+/-} mice, we used a farnesyl-transferase inhibitor (FTI; BMS 191563; ref. 16). This agent reduces Ras signalling by blocking farnesylation, a post-translational modification that is essential for Ras function¹⁷. FTIs have been shown to rescue other NF1 phenotypes, such as Schwann cell proliferation in human and murine cells^{18,19}. Wild-type and *Nf1*^{+/-} mice were injected intraperitoneally every day 60 min before training with either saline or 5 mg per kg (body weight) BMS 191563; the training conditions and the genetic background of the animals were the same as in the *Nf1*^{+/-}/*K-ras*^{+/-} experiment.

Analysis of the day 7 probe trial showed an effect of genotype—

† Present address: Memory Pharmaceuticals Corporation, 100 Philips Parkway, Montvale, New Jersey 07645, USA.

treatment interaction on searching strategy ($F_{3,32} = 2.83$, $P < 0.05$; Fig. 1h). $Nf1^{+/-}$ mice treated with BMS191563 searched significantly longer in the training quadrant than $Nf1^{+/-}$ mice treated with saline ($P < 0.05$), and were indistinguishable from wild-type mice ($P > 0.05$). Moreover, analysis of proximity scores showed that $Nf1^{+/-}$ mice treated with BMS191563 searched selectively ($t_{18} = 4.13$, $P < 0.05$; Fig. 1i), as did wild-type mice ($t_{17} = 4.33$, $P < 0.05$), whereas $Nf1^{+/-}$ mice treated with saline did not ($t_{19} = 0.266$, $P > 0.05$). Administration of 5 mg per kg BMS191563 did not adversely affect wild-type animals, as these mice searched equivalently to wild-type mice treated with saline ($P > 0.05$; Fig. 1h). The finding that the learning deficits of $Nf1^{+/-}$ mice can be reversed with an FTI confirms our genetic results, and shows that these learning deficits are reversible in adult mice.

Learning is thought to occur through activity-dependent synaptic modifications in neuronal networks²⁰. To determine whether changes in synaptic plasticity might account for the spatial (hippocampal-dependent⁹) learning impairments of $Nf1^{+/-}$ mutants, we examined long-term potentiation (LTP) at Schaffer collateral/CA1 synapses in hippocampal slices from these mice. Schaffer collateral stimulation (60 μ A) elicited a similar baseline excitatory post-synaptic potential (EPSP) in $Nf1^{+/-}$ and wild-type mice ($Nf1^{+/-} = 1.02 \pm 0.06$ mV, wild type = 1.09 ± 0.05 , $F_{1,54} = 0.58$, $P > 0.05$).

We first induced LTP using a theta-burst stimulation (TBS) protocol, which mimics the *in vivo* activity of hippocampal neurons

during exploratory behaviour²¹. This protocol consisted of two, five or ten bursts. The two-burst protocol elicited a persistent potentiation in wild-type animals (Fig. 2a), which was significantly larger than that triggered in $Nf1^{+/-}$ mice ($F_{1,11} = 6.98$, $P < 0.05$). This difference was also seen with five ($F_{1,13} = 8.27$, $P < 0.05$; Fig. 2b) and ten bursts ($F_{1,16} = 6.93$, $P < 0.05$; Fig. 2c). Thus, TBS stimulation shows that there is an LTP deficit in $Nf1^{+/-}$ mice (Fig. 2e). We also examined LTP induced under high-frequency stimulation (HFS, 100 Hz for 1 s; Fig. 2e). With this protocol, the potentiation induced was equivalent in wild-type and $Nf1^{+/-}$ mice ($F_{1,15} = 0.06$, $P > 0.05$; see below).

To determine whether the TBS-induced LTP deficits in $Nf1^{+/-}$ mice were caused by increased Ras function, we tested $Nf1^{+/-}/K-ras^{+/-}$ mice (of the same genetic background as those tested in the water maze). The magnitude of CA1 LTP induced with a two-burst protocol differed across genotypes ($F_{3,35} = 4.6$, $P < 0.05$; Fig. 2d). The amount of LTP induced in $Nf1^{+/-}/K-ras^{+/-}$ was significantly higher than that induced in $Nf1^{+/-}$ mice (Fisher PLSD, $P < 0.05$), and equivalent to that of wild-type mice ($P > 0.05$; Fig. 2f). $K-ras^{+/-}$ mice were also significantly impaired relative to both $Nf1^{+/-}/K-ras^{+/-}$ and wild-type mice (LTP at 40 min 114 ± 4.47 , $P < 0.05$; data not shown). These results show that the $K-ras^{+/-}$ mutation can rescue the LTP deficits in $Nf1^{+/-}$ mice, just as it can also rescue their spatial learning deficits.

$Nf1^{+/-}$ mice also showed a decrease in the input–output function

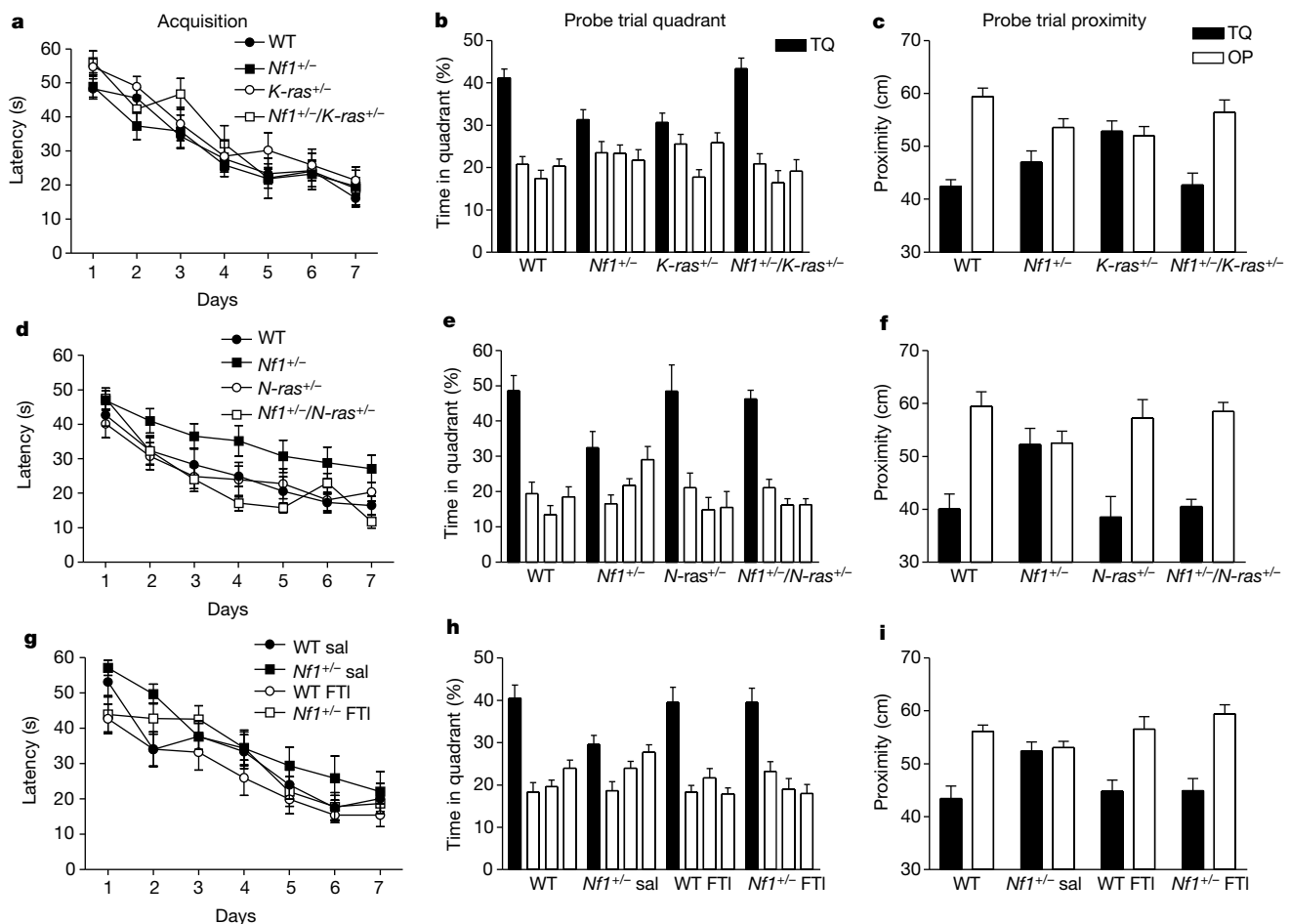


Figure 1 Learning deficits of $Nf1^{+/-}$ mice are Ras dependent. **a–c**, Hidden version of the water maze for the $Nf1^{+/-}/K-ras^{+/-}$ population (wild type (WT), $n = 24$; $Nf1^{+/-}$, $n = 15$; $K-ras^{+/-}$, $n = 15$; $Nf1^{+/-}/K-ras^{+/-}$, $n = 11$). **a**, Latency to get to the platform over days. **b**, Per cent time spent in each quadrant during a probe trial **c**, Average proximity to the exact position where the platform was during training, compared with proximity to the opposite position in the pool. **d–f**, Acquisition, per cent time in quadrant, and proximity

data for the $Nf1^{+/-}/N-ras^{+/-}$ population (WT, $n = 10$; $Nf1^{+/-}$, $n = 10$; $N-ras^{+/-}$, $n = 7$; $Nf1^{+/-}/N-ras^{+/-}$, $n = 9$). **g–i**, Acquisition, per cent time in quadrant, and proximity data for the different genotypes and treatments during the FTI rescue experiment (WT_{FTI}, $n = 19$; WT_{saline}, $n = 18$; $Nf1^{+/-}$ _{FTI}, $n = 18$; $Nf1^{+/-}_{saline}, $n = 18$). Quadrants are training quadrant (TQ), adjacent right, adjacent left and opposite quadrant (OP).$

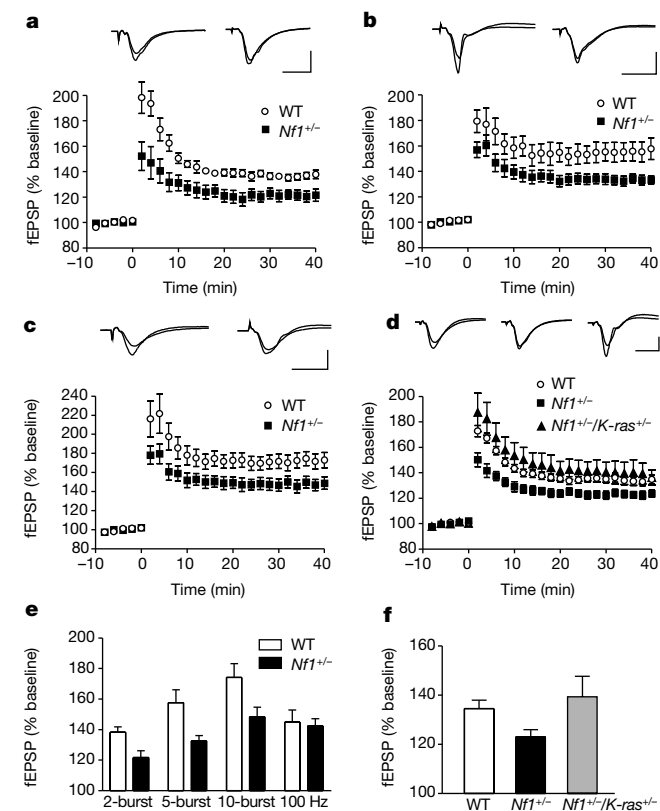


Figure 2 Ras-dependent long-term potentiation deficits in *Nf1*^{+/-} animals. Percentage of baseline field EPSP (fEPSP) is plotted over time. **a**, Two-burst induction protocol (wild type (WT), *n* = 5; *Nf1*^{+/-}, *n* = 8). **b**, Five-burst induction protocol (WT, *n* = 7; *Nf1*^{+/-}, *n* = 8). **c**, Ten-burst induction protocol (WT, *n* = 7; *Nf1*^{+/-}, *n* = 11). **d**, Two-burst induction protocol. LTP deficits in *Nf1*^{+/-} mice are Ras dependent (WT, *n* = 13; *Nf1*^{+/-}, *n* = 13; *Nf1*^{+/-}/*K-ras*^{+/-}, *n* = 8). Representative traces are shown from left to right for WT, *Nf1*^{+/-} and *Nf1*^{+/-}/*K-ras*^{+/-}. Horizontal bar, 10 ms; vertical bar, 1 mV. **e**, LTP measured 40 min after induction under different stimulation protocols. **f**, LTP measured 40 min after induction in the *Nf1*^{+/-}/*K-ras*^{+/-} rescue experiment.

of extracellularly recorded EPSPs at the Schaffer collateral/CA1 synapse ($F_{1,16} = 4.11$, $P < 0.05$, Fisher PLSD (100 μ A), $P < 0.05$; Fig. 3a). Neither this impairment, nor the LTP deficit, seemed to be caused by abnormal presynaptic function, because electrophysiological phenomena sensitive to presynaptic changes such as paired-pulse facilitation ($F_{1,19} = 0.43$, $P > 0.05$; Fig. 3b), augmentation, depletion and probability of glutamate release²² were normal in *Nf1*^{+/-} mice (data not shown). The LTP deficits of *Nf1*^{+/-} mice were revealed by TBS, but not HFS stimulation. LTP induced by TBS is more sensitive to changes in GABA-mediated inhibition than is LTP induced by HFS²³. In addition, Ras can affect chloride currents in adrenal gland cells²⁴. We therefore investigated whether GABA-mediated inhibition was altered in *Nf1*^{+/-} mice. We measured evoked inhibitory postsynaptic potentials (IPSPs) in CA1 pyramidal cells of hippocampal slices from *Nf1*^{+/-} mice. We found that *Nf1*^{+/-} mice have larger IPSPs than those of wild-type controls (ANOVA, $F_{1,88} = 5.31$, $P < 0.05$; Fisher PLSD, $P < 0.05$; Fig. 3c). We also measured monosynaptically evoked IPSPs in the presence of AP5 (2-amino-5-phosphonopentanoic acid), an NMDA (*N*-methyl-D-aspartate) receptor antagonist, and CNQX (6-cyano-7-nitroquinoxaline-2,3-dione), an AMPA (α -amino-3-hydroxy-5-methylisoxazole-4-propionic acid) receptor antagonist (Fig. 3d). Once again, IPSPs were larger in *Nf1*^{+/-} than in wild-type mice (ANOVA, $F_{2,237} = 5.09$, $P < 0.05$; Fisher PLSD, $P < 0.05$; Fig. 3d). The increased inhibition in *Nf1*^{+/-} mice was also rescued by manipulations that decrease Ras function: IPSPs in *Nf1*^{+/-}/*K-ras*^{+/-}

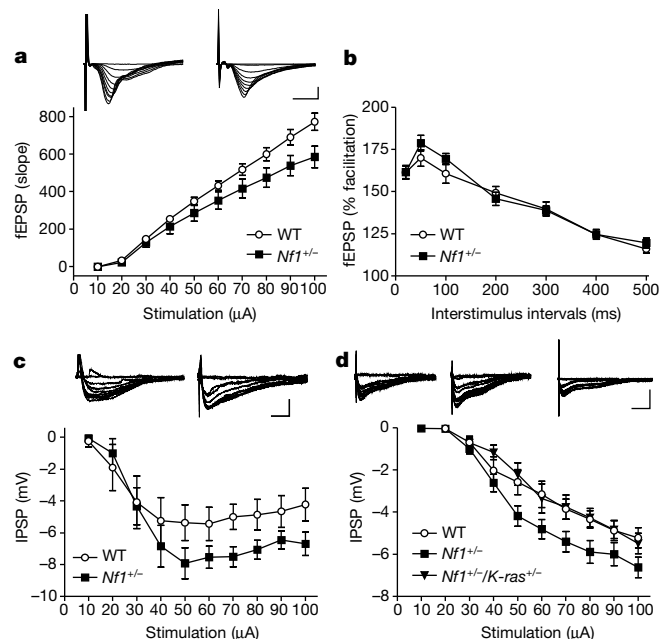


Figure 3 Ras-dependent enhanced inhibition in *Nf1*^{+/-} mice. **a**, Input-output function at the CA3–CA1 synapse (wild type (WT), *n* = 9; *Nf1*^{+/-}, *n* = 9) with representative traces (left, WT; right, *Nf1*^{+/-}). Horizontal bar, 5 ms; vertical bar, 0.5 mV. **b**, Paired-pulse facilitation (WT, *n* = 9; *Nf1*^{+/-}, *n* = 12) at different interstimulus intervals. **c**, Evoked IPSPs measured intracellularly in CA1 pyramidal cells using different stimulation strengths (WT, 5 mice, 13 neurons, *Nf1*^{+/-}, 4 mice, 17 neurons). **d**, Monosynaptically evoked IPSPs measured in the presence of AP5 and CNQX (WT, 7 mice, 17 neurons; *Nf1*^{+/-}, 11 mice, 23 neurons; *Nf1*^{+/-}/*K-ras*^{+/-}, 6 mice, 13 neurons). **e**, **f**, Representative traces are shown left to right for WT, *Nf1*^{+/-} and *Nf1*^{+/-}/*K-ras*^{+/-} mice. Horizontal bar, 100 ms; vertical bar, 5 mV.

mice are smaller than in *Nf1*^{+/-} mice ($P < 0.05$; Fig. 3d) and indistinguishable from wild-type controls ($P > 0.05$).

Results from several of our experiments suggest that this increase in GABA-mediated inhibition is the cause for the LTP deficits in *Nf1*^{+/-} mice. First, picrotoxin (10 μ M), a GABA_A receptor antagonist, rescues the LTP (two-burst) deficits of *Nf1*^{+/-} mice ($F_{1,9} = 0.78$, $P > 0.05$; Fig. 4a). Second, LTP deficits are evident in *Nf1*^{+/-} mice (Fig. 2a) at synaptic stimulation strengths (60 μ A) that cause an increase in GABA-mediated inhibition, but are absent at synaptic stimulation strengths (35 μ A) that do not cause increases in inhibition ($F_{1,27} = 0.263$, $P > 0.05$; Figs 3c and 4b). Note that in wild-type mice the LTP induced at a stimulation strength of 60 μ A is larger than that induced at 35 μ A ($F_{1,16} = 18.1$, $P < 0.05$), whereas in *Nf1*^{+/-} mice it is not ($F_{1,22} = 0.18$, $P > 0.05$; Fig. 4c). This is consistent with the larger inhibition in *Nf1*^{+/-} mice at 60 μ A (Fig. 4c). Third, LTP induced with HFS, which is less sensitive to inhibition than LTP induced with TBS, is normal in *Nf1*^{+/-} mice (Fig. 2e).

Our results show that the spatial learning impairments in a mouse model of NF1 can be rescued by two different genetic manipulations that decrease Ras levels. Our data also show that the learning deficits in *Nf1*^{+/-} mice can be reversed pharmacologically in the adult—an important finding for the development of treatments for learning deficits associated with NF1. Our results suggest that abnormally high or low Ras activity can disrupt learning, indicating that precise Ras modulation by neurofibromin is essential for learning and memory. Although the learning deficits in *Drosophila* that lack neurofibromin are dependent on adenylyl cyclase⁶, *Nf1*^{+/-} mice have normal adenylyl cyclase activity²⁵. These mice, however, have upregulated Ras/mitogen-activated protein kinase activity²⁶. Our data indicate that increased Ras activity,

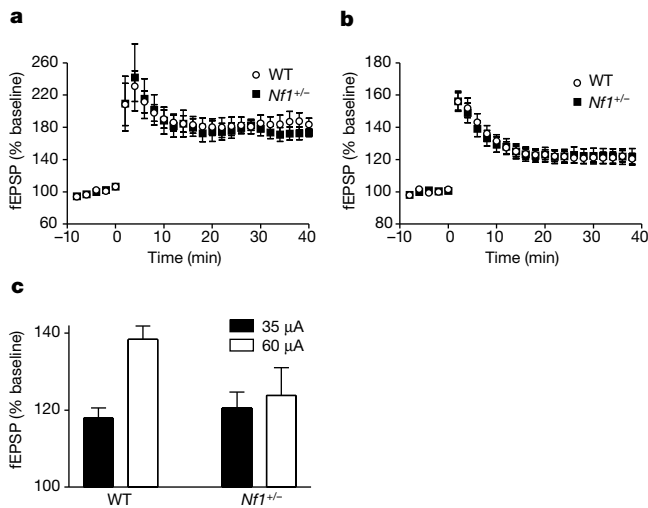


Figure 4 Long-term potentiation deficits in *Nf1*^{+/-} mice are caused by increased inhibition. **a**, LTP induction under picrotoxin using a two-burst induction (wild type (WT), *n* = 6; *Nf1*^{+/-}, *n* = 5). **b**, LTP induced with two bursts using a stimulation strength of 35 μ A (evokes similar IPSPs in WT and *Nf1*^{+/-} mice, see Fig. 3c) is not different in WT (*n* = 13) and *Nf1*^{+/-} (*n* = 16) mice. **c**, In WT mice the LTP induced at a stimulation strength of 60 μ A is larger than LTP induced at 35 μ A, whereas in *Nf1*^{+/-} mice it is not (60 μ A evokes larger IPSPs in *Nf1*^{+/-} mice).

caused by partial loss of neurofibromin, leads to abnormally high GABA-mediated inhibition, and that this enhancement in inhibition may be responsible for impairments in synaptic plasticity that might underlie the learning deficits associated with NF1. These findings suggest that strategies that decrease either Ras activity or GABA-mediated inhibition might be used to treat the learning deficits associated with NF1. □

Methods

Animals

Generation of the different genetically modified mice has been described^{12,27,28}. For the *Nf1*^{+/-}/*N-ras*^{+/-} experiment, we generated animals by crossing *Nf1*^{+/-} mice, which had been previously backcrossed five generations to the C57BL/6N background from 129T2/SvEmsJ (N5 into C57BL/6N), with *N-ras*^{+/-} also backcrossed five generations to the C57BL/6N background. All the other animals were 129T2/SvEmsJ–C57BL/6N F₁ hybrids generated by crossing *Nf1*^{+/-} mice (backcrossed more than 11 generations to the C57BL/6N background) with wild-type or *K-ras*^{+/-} mice on the 129T2/SvEmsJ background. Thus, in every experiment the mice from each genotype were littermates and of isogenic genetic background. All experiments were carried out blind with respect to genotype or treatment.

Water maze

The basic protocol for the water maze experiments has been described¹⁰. Mice from the 129T2/SvEmsJ–C57BL/6N F₁ background were given two trials per day (30-s intertrial intervals) with a probe trial (60 s) at the end of training day 7. Because mice from the C57BL/6N genetic background take longer to learn the water maze²⁹, mice from the *Nf1*^{+/-}/*N-ras*^{+/-} population (N5 into C57BL/6N) were given three blocks of two trials per day, with a probe trial at training day 5. We used contextual fear conditioning in wild-type mice (129T2/SvEmsJ–C57BL/6N F₁) to determine the dose used in the BMS191563 experiments. With one single injection 6 h before training, a dose of 10 mg per kg (body mass) did not affect learning but higher dosages did (time freezing: saline, 63.3%; 10 mg per kg, 55.8%; 30 mg per kg, 34.4%; 45 mg per kg, 46.7%). But 10 mg per kg injected for 7 days (as in the water maze experiment) affected learning in wild-type mice (saline, 82.9%; 10 mg per kg, 52.2%; *F*_{1,10} = 9.43, *P* < 0.05). As shown, 5 mg per kg injected for 7 days did not affect learning in wild-type mice. We dissolved BMS191563 in sterile saline solution and injected it every day 1 h before the experiment started.

Electrophysiology

For the field potentials, recordings were made from transverse hippocampal slices (400 μ m thick) in a submerged recording chamber perfused (2 ml min⁻¹) with artificial cerebrospinal fluid (ACSF) containing (in mM): 120 NaCl, 3.5 KCl, 2.5 CaCl₂, 1.3 MgSO₄, 1.25 NaH₂PO₄, 26 NaHCO₃ and 10 D-glucose at 30 °C (saturated with 95% O₂ and 5% CO₂). For the LTP experiments, EPSPs were evoked alternatively in separate pathways (control and tetanized) in the CA1 Schaffer collateral/commissural afferents with 100- μ s test pulses through two stimulating electrodes (~300 μ m from the Pt/Ir recording

electrode). Unless indicated otherwise, the stimulation strength in both stimulating electrodes was set to 60 μ A. After a 10-min baseline period, LTP was induced in one pathway according to a HFS (100 Hz for 1 s) or TBS protocol (two, five or ten bursts, each burst four pulses at 100 Hz, 200-ms interburst intervals). We excluded from further analysis slices in which there was significant drift in the control pathway or in which seizures developed. The amount of potentiation was calculated as a percentage of the baseline EPSP slope. For the experiments with picrotoxin, the drug was dissolved (10 μ M) in ACSF and was present throughout the experiment. For the input–output curves, we applied different stimulation strengths (from 10 to 100 μ A in steps of 10 μ A). For each slice, five EPSPs were averaged per stimulation strength. Paired-pulse facilitation was assessed by measuring the per cent increase in the slope of the second pulse in relation to the first (20-, 50-, 100-, 200-, 300-, 400-, 500-ms interpulse intervals). The stimulation strength was adjusted to one-third of the maximum EPSP amplitude.

To assess inhibition in *Nf1*^{+/-} mice, we measured IPSPs from CA1 pyramidal neurons using whole-cell (blind technique³⁰) bridge mode recordings (Axoclamp 2B, Axon Instruments). IPSPs were evoked through a stimulating electrode placed in the Schaffer collateral/commissural afferents (~75 μ m from the recording pipette when AP5 and CNQX were applied) that applied different stimulation strengths (from 10 to 100 μ A in steps of 10 μ A). The IPSP amplitude was measured, with five IPSPs averaged for each neuron per stimulation strength. The intracellular solution contained (in mM): 135 potassium gluconate, 5 HEPES, 2 Mg²⁺-ATP, 5 MgCl₂, 0.3 GTP, 0.05 EGTA (pipette resistance 7–11 M Ω). To evoke IPSPs monosynaptically, AP5 and CNQX (10 μ M) were present in the ACSF.

In all experiments, data were averaged and entered into analysis as a single subject, and therefore reflect individual mice rather than individual slices or neurons.

Statistical analysis

We analysed acquisition data from the water maze by repeated-measures ANOVA. Per cent time in training quadrant for the different genotypes was analysed using single-factor ANOVA; post-hoc comparisons (Fisher's PLSD) between genotypes were carried out when appropriate. Planned comparisons using a paired *t*-test were used to analyse the proximity data. LTP was analysed using single-factor ANOVA on the average amount of LTP 30–40 min after induction. We analysed the inhibition and input–output curves using ANOVA, and post-hoc comparisons were performed when appropriate.

Received 10 August; accepted 21 November 2001.

Published online 16 January 2002, DOI 10.1038/nature7111.

- North, K. Neurofibromatosis type 1. *Am. J. Med. Genet.* **97**, 119–127 (2000).
- Silva, A. J. et al. A mouse model for the learning and memory deficits associated with neurofibromatosis type I. *Nature Genet.* **15**, 281–284 (1997).
- Ballester, R. et al. The *Nf1* locus encodes a protein functionally related to mammalian GAP and yeast IRA proteins. *Cell* **63**, 851–859 (1990).
- Xu, G. F. et al. The catalytic domain of the neurofibromatosis type 1 gene product stimulates Ras GTPase and complements *ira* mutants of *S. cerevisiae*. *Cell* **63**, 835–841 (1990).
- Guo, H. F., The, L., Hannan, F., Bernards, A. & Zhong, Y. Requirement of *Drosophila* NF1 for activation of adenyl cyclase by PACAP38-like neuropeptides. *Science* **276**, 795–798 (1997).
- Guo, H. F., Tong, J., Hannan, F., Luo, L. & Zhong, Y. A neurofibromatosis-1-regulated pathway is required for learning in *Drosophila*. *Nature* **403**, 895–898 (2000).
- Xu, H. & Gutmann, D. H. Mutations in the GAP-related domain impair the ability of neurofibromin to associate with microtubules. *Brain Res.* **759**, 149–152 (1997).
- Ozonoff, S. Cognitive impairment in neurofibromatosis type 1. *Am. J. Med. Genet.* **89**, 45–52 (1999).
- Morris, R. G., Garrud, P., Rawlins, J. N. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **297**, 681–683 (1982).
- Costa, R. M. et al. Learning deficits, but normal development and tumour predisposition, in mice lacking exon 23a of *Nf1*. *Nature Genet.* **27**, 399–405 (2001).
- Klose, A. et al. Selective disactivation of neurofibromin GAP activity in neurofibromatosis type 1. *Hum. Mol. Genet.* **7**, 1261–1268 (1998).
- Johnson, L. K. et al. *ras* is an essential gene in the mouse with partial functional overlap with *N-ras*. *Genes Dev.* **11**, 2468–2481 (1997).
- Brannan, C. I. et al. Targeted disruption of the neurofibromatosis type-1 gene leads to developmental abnormalities in heart and various neural crest-derived tissues. *Genes Dev.* **8**, 1019–1029 (1994).
- Brandeis, R., Brandys, Y. & Yehuda, S. The use of the Morris water maze in the study of memory and learning. *Int. J. Neurosci.* **48**, 29–69 (1989).
- Gallagher, M., Burwell, R. & Burchinal, M. Severity of spatial learning impairment in aging: Development of a learning index for performance in the Morris water maze. *Behav. Neurosci.* **107**, 618–626 (1993).
- Muthalif, M. M. et al. Contribution of Ras GTPase/MAP kinase and cytochrome P450 metabolites to deoxycorticosterone-salt-induced hypertension. *Hypertension* **35**, 457–463 (2000).
- Gibbs, J. B. et al. Farnesyltransferase inhibitors versus Ras inhibitors. *Curr. Opin. Chem. Biol.* **1**, 197–203 (1997).
- Yan, N. et al. Farnesyltransferase inhibitors block the neurofibromatosis type I (NF1) malignant phenotype. *Cancer Res.* **55**, 3569–3575 (1995).
- Kim, H. A., Ling, B. & Ratner, N. *Nf1*-deficient mouse Schwann cells are angiogenic and invasive and can be induced to hyperproliferate: reversion of some phenotypes by an inhibitor of farnesyl protein transferase. *Mol. Cell. Biol.* **17**, 862–872 (1997).
- Abbott, L. F. & Nelson, S. B. Synaptic plasticity: taming the beast. *Nature Neurosci.* **3**, 1178–1183 (2000).
- Larson, J., Wong, D. & Lynch, G. Patterned stimulation at the theta frequency is optimal for the induction of hippocampal long-term potentiation. *Brain Res.* **368**, 347–350 (1986).
- Hessler, N. A., Shirke, A. M. & Malinow, R. The probability of transmitter release at a mammalian central synapse. *Nature* **366**, 569–572 (1993).
- Chapman, C. A., Perez, Y. & Lacaille, J. C. Effects of GABA inhibition on the expression of long-term potentiation in CA1 pyramidal cells are dependent on tetanization parameters. *Hippocampus* **8**, 289–298 (1998).

24. Chorvatova, A., Gendron, L., Bilodeau, L., Gallo-Payet, N. & Payet, M. D. A Ras-dependent chloride current activated by adrenocorticotropin in rat adrenal zona glomerulosa cells. *Endocrinology* **141**, 684–692 (2000).
25. Tong, J. *et al.* NF1-regulated adenylyl cyclase pathway. *Soc. Neurosci. Abstr.* abstract no. 345.9 (Society for Neuroscience, New Orleans, 2000).
26. Ingram, D. A. *et al.* Hyperactivation of p21(ras) and the hematopoietic-specific Rho GTPase, Rac2, cooperate to alter the proliferation of neurofibromin-deficient mast cells in vivo and in vitro. *J. Exp. Med.* **194**, 57–69 (2001).
27. Jacks, T. *et al.* Tumour predisposition in mice heterozygous for a targeted mutation in *Nf1*. *Nature Genet.* **7**, 353–361 (1994).
28. Umanoff, H., Edelmann, W., Pellicci, A. & Kucherlapati, R. The murine *N-ras* gene is not essential for growth and development. *Proc. Natl Acad. Sci. USA* **92**, 1709–1713 (1995).
29. Voikar, V., Koks, S., Vasar, E. & Rauvala, H. Strain and gender differences in the behaviour of mouse lines commonly used in transgenic studies. *Physiol. Behav.* **72**, 271–281 (2001).
30. Blanton, M. G., Lo Turco, J. J. & Kriegstein, A. R. Whole cell recording from neurons in slices of reptilian and mammalian cerebral cortex. *J. Neurosci. Methods* **30**, 203–210 (1989).

Acknowledgements

We thank V. Manne for the BMS191563, and E. Friedman for technical assistance in earlier experiments. We are grateful to M. Barad, D. Buonomano, T. Cannon, J. Colicelli, P. Frankland, L. Kaczmarek, A. Matynia, M. Sanders and D. Smith for discussions, and to C. Brannan and S. Schlussel for encouragement. R.M.C. received support from the Graduated Program in Basic and Applied Biology (GABBA) of the University of Oporto, the Portuguese Foundation for Science and Technology (FCT) and the National Neurofibromatosis Foundation (NNF). This work was also supported by a generous donation from K. M. Spivak, and by grants from the NIH (R01 NS38480), Neurofibromatosis Inc. (National, Illinois, Mass Bay Area, Minnesota, Arizona, Kansas and Central Plains, Mid-Atlantic, and Texas chapters), the Merck and the NNF foundations to A.J.S.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.J.S. (e-mail: Silvaa@mednet.ucla.edu).

Gene expression profiling predicts clinical outcome of breast cancer

Laura J. van 't Veer^{*,†}, Hongyue Dai^{†,‡}, Marc J. van de Vijver^{*,†}, Yudong D. He[‡], Augustinus A. M. Hart^{*}, Mao Mao[‡], Hans L. Peterse^{*}, Karin van der Kooy^{*}, Matthew J. Marton[‡], Anke T. Witteveen^{*}, George J. Schreiber[‡], Ron M. Kerkhoven^{*}, Chris Roberts[‡], Peter S. Linsley[‡], René Bernards^{*} & Stephen H. Friend[‡]

^{*} Divisions of Diagnostic Oncology, Radiotherapy and Molecular Carcinogenesis and Center for Biomedical Genetics, The Netherlands Cancer Institute, 121 Plesmanlaan, 1066 CX Amsterdam, The Netherlands

[‡] Rosetta Inpharmatics, 12040 115th Avenue NE, Kirkland, Washington 98034, USA

[†] These authors contributed equally to this work

Breast cancer patients with the same stage of disease can have markedly different treatment responses and overall outcome. The strongest predictors for metastases (for example, lymph node status and histological grade) fail to classify accurately breast tumours according to their clinical behaviour^{1–3}. Chemotherapy or hormonal therapy reduces the risk of distant metastases by approximately one-third; however, 70–80% of patients receiving this treatment would have survived without it^{4,5}. None of the signatures of breast cancer gene expression reported to date^{6–12} allow for patient-tailored therapy strategies. Here we used DNA microarray analysis on primary breast tumours of 117 young patients, and applied supervised classification to identify a gene expression signature strongly predictive of a short interval to distant metastases ('poor prognosis' signature) in patients without tumour cells in local lymph nodes at diagnosis (lymph node negative). In addition, we established a signature that identifies tumours of *BRCA1* carriers. The poor prognosis signature consists of genes regulating cell cycle, invasion, metastasis and

angiogenesis. This gene expression profile will outperform all currently used clinical parameters in predicting disease outcome. Our findings provide a strategy to select patients who would benefit from adjuvant therapy.

We selected 98 primary breast cancers: 34 from patients who developed distant metastases within 5 years, 44 from patients who continued to be disease-free after a period of at least 5 years, 18 from patients with *BRCA1* germline mutations, and 2 from *BRCA2* carriers. All 'sporadic' patients were lymph node negative, and under 55 years of age at diagnosis. From each patient, 5 µg total RNA was isolated from snap-frozen tumour material and used to derive complementary RNA (cRNA). A reference cRNA pool was made by pooling equal amounts of cRNA from each of the sporadic carcinomas. Two hybridizations were carried out for each tumour using a fluorescent dye reversal technique on microarrays containing approximately 25,000 human genes synthesized by inkjet technology¹³. Fluorescence intensities of scanned images were quantified, normalized and corrected to yield the transcript abundance of a gene as an intensity ratio with respect to that of the signal of the reference pool¹⁴. Some 5,000 genes were significantly regulated across the group of samples (that is, at least a twofold difference and a *P*-value of less than 0.01 in more than five tumours).

An unsupervised, hierarchical clustering algorithm allowed us to cluster the 98 tumours on the basis of their similarities measured over these approximately 5,000 significant genes. Similarly, the ~5,000 genes were clustered on the basis of their similarities measured over the group of 98 tumours (Fig. 1a). In the dendrograms shown in Fig. 1a (left and top), the length and the subdivision of the branches displays the relatedness of the breast tumours (left) and the expression of the genes (top). Two distinct groups of tumours are the dominant feature in this two-dimensional display (top and bottom of plot, representing 62 and 36 tumours, respectively), suggesting that the tumours can be divided into two types on the basis of this set of ~5,000 significant genes. Notably, in the upper group only 34% of the sporadic patients were from the group who developed distant metastases within 5 years, whereas in the lower group 70% of the sporadic patients had progressive disease (Fig. 1b). Thus, using unsupervised clustering we can already, to some extent, distinguish between 'good prognosis' and 'poor prognosis' tumours.

To gain insight into the genes of the dominant expression signatures, we associated them with histopathological data; for example, oestrogen receptor (ER)-α expression as determined by immunohistochemical (IHC) staining (Fig. 1b). Out of 39 IHC-stained tumours negative for ER-α expression (ER negative), 34 clustered together in the bottom branch of the tumour dendrogram. In the enlargement shown in Fig. 1c, a group of downregulated genes is represented containing both the ER-α gene (*ESR1*) and genes that are apparently co-regulated with ER, some of which are known ER target genes. A second dominant gene cluster is associated with lymphocytic infiltrate and includes several genes expressed primarily by B and T cells (Fig. 1d).

Sixteen out of eighteen tumours of *BRCA1* carriers are found in the bottom branch intermingled with sporadic tumours. This is consistent with the idea that most *BRCA1* mutant tumours are ER negative and manifest a higher amount of lymphocytic infiltrate¹⁵. The two tumours of *BRCA2* carriers are part of the upper cluster of tumours and do not show similarity with *BRCA1* tumours. Neither high histological grade nor angiogenesis is a specific feature of either of the clusters (Fig. 1b). We conclude that unsupervised clustering detects two subgroups of breast cancers, which differ in ER status and lymphocytic infiltration. A similar conclusion has also been reported previously^{7,16}.

The 78 sporadic lymph-node-negative patients were selected specifically to search for a prognostic signature in their gene expression profiles. Forty-four patients remained free of disease

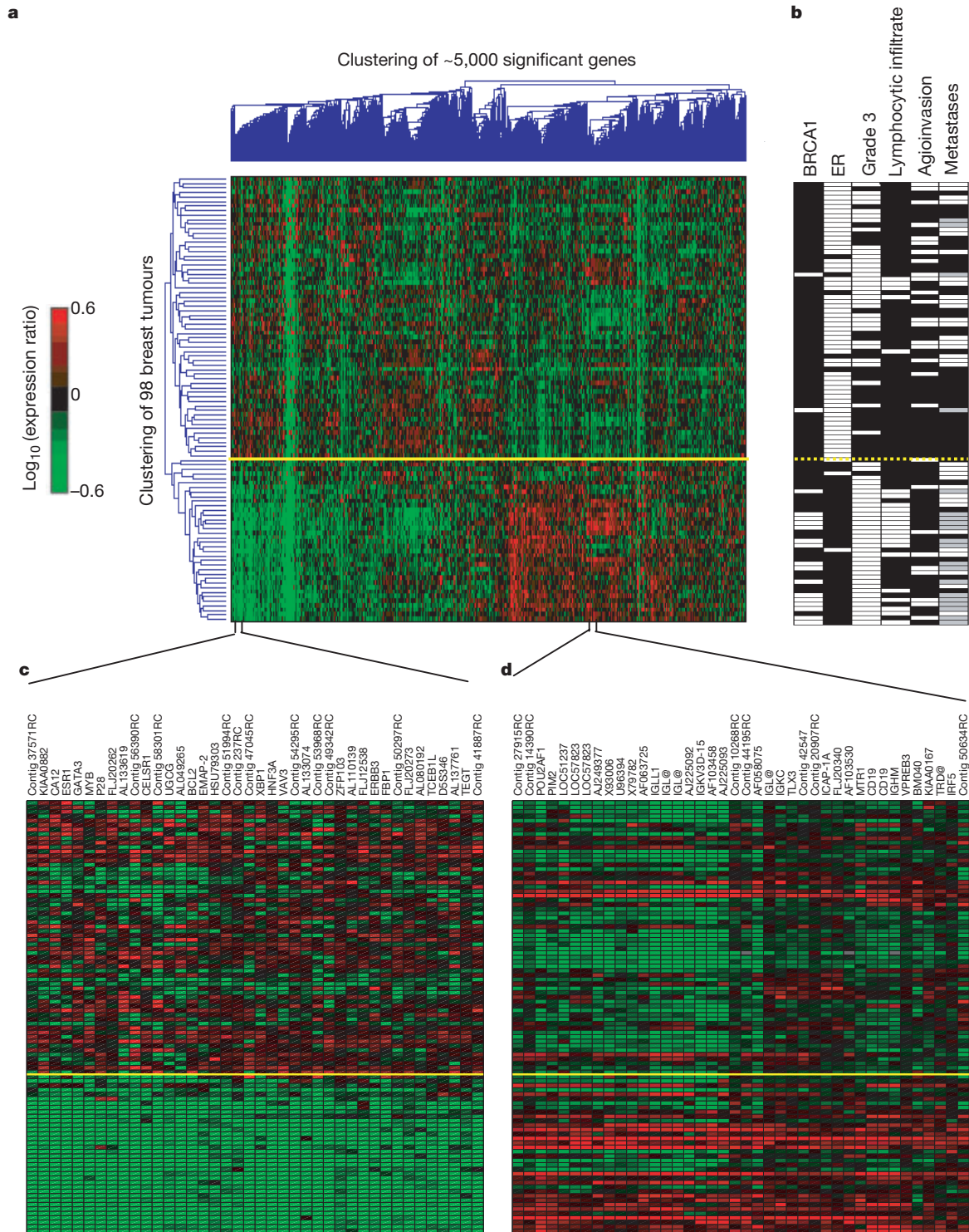


Figure 1 Unsupervised two-dimensional cluster analysis of 98 breast tumours. **a**, Two-dimensional presentation of transcript ratios for 98 breast tumours. There were 4,968 significant genes across the group. Each row represents a tumour and each column a single gene. As shown in the colour bar, red indicates upregulation, green downregulation, black no change, and grey no data available. The yellow line marks the subdivision into two dominant tumour clusters. **b**, Selected clinical data for the 98 patients in **a**: *BRCA1* germline mutation carrier (or sporadic patient), ER expression, tumour grade 3 (versus grade 1 and 2), lymphocytic infiltrate, angioinvasion, and metastasis status. White indicates positive, black negative and grey denotes tumours derived from *BRCA1*

germline carriers who were excluded from the metastasis evaluation. The cluster below the yellow line consists of 36 tumours, of which 34 are ER negative (total 39 ER-negative) and 16 are carriers of the *BRCA1* mutation (total 18). **c**, Enlarged portion from **a** containing a group of genes that co-regulate with the ER- α gene (*ESR1*). Each gene is labelled by its gene name or accession number from GenBank. Contig ESTs ending with RC are reverse-complementary of the named contig EST. **d**, Enlarged portion from **a** containing a group of co-regulated genes that are the molecular reflection of extensive lymphocytic infiltrate, and comprise a set of genes expressed in T and B cells. (Gene annotation as in **c**.)

after their initial diagnosis for an interval of at least 5 years (good prognosis group, mean follow-up of 8.7 years), and 34 patients had developed distant metastases within 5 years (poor prognosis group, mean time to metastases 2.5 years) (Fig. 2a). To identify reliably good and poor prognostic tumours, we used a powerful three-step supervised classification method, similar to those used previously^{8,17,18}. In brief, approximately 5,000 genes (significantly regulated in more than 3 tumours out of 78) were selected from the

25,000 genes on the microarray. The correlation coefficient of the expression for each gene with disease outcome was calculated and 231 genes were found to be significantly associated with disease outcome (correlation coefficient <-0.3 or >0.3). In the second step, these 231 genes were rank-ordered on the basis of the magnitude of the correlation coefficient. Third, the number of genes in the 'prognosis classifier' was optimized by sequentially adding subsets of 5 genes from the top of this rank-ordered list and

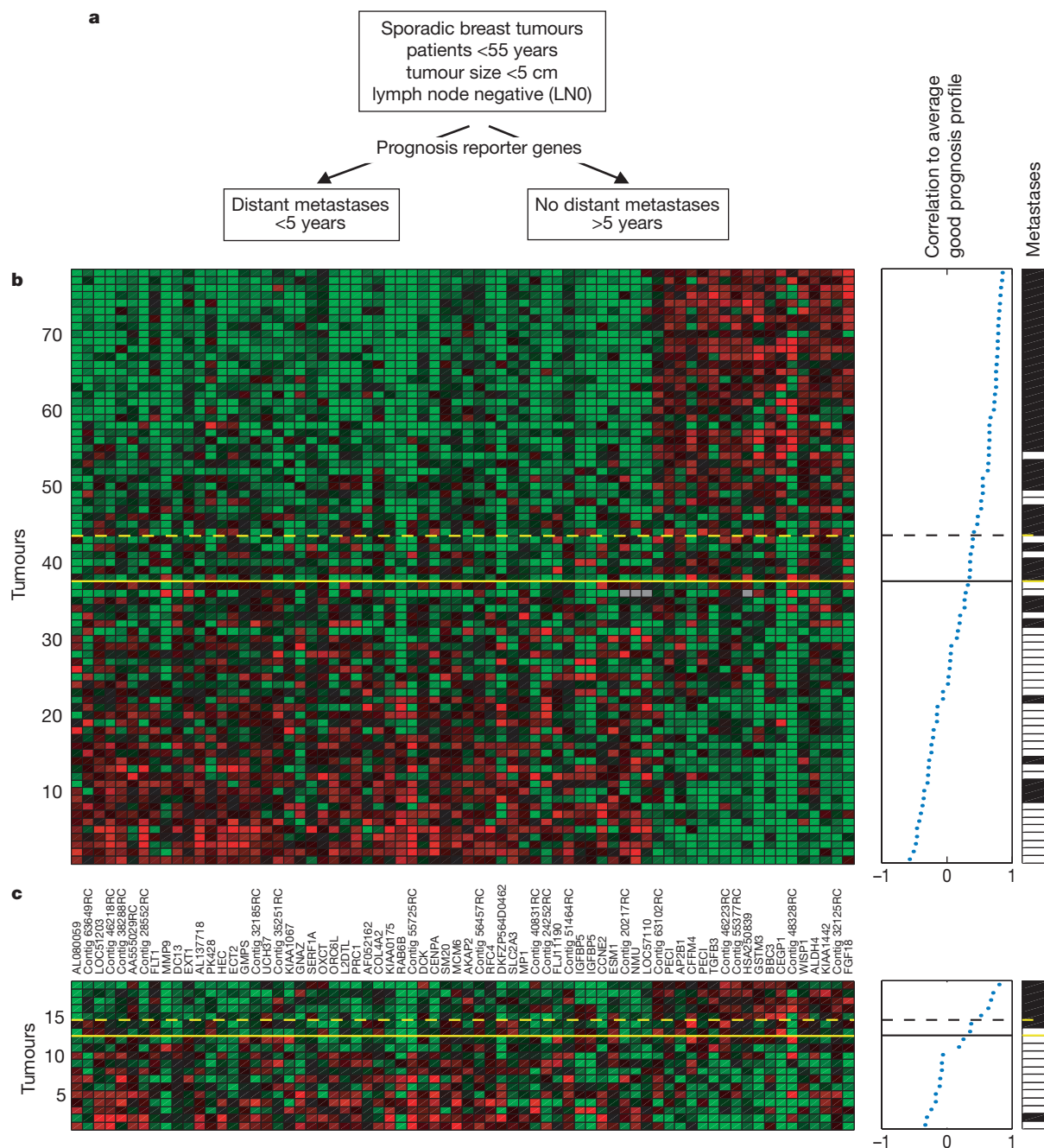


Figure 2 Supervised classification on prognosis signatures. **a**, Use of prognostic reporter genes to identify optimally two types of disease outcome from 78 sporadic breast tumours into a poor prognosis and good prognosis group (for patient data see Supplementary Information Table S1). **b**, Expression data matrix of 70 prognostic marker genes from tumours of 78 breast cancer patients (left panel). Each row represents a tumour and each column a gene, whose name is labelled between **b** and **c**. Genes are ordered according to their correlation coefficient with the two prognostic groups. Tumours are ordered by the correlation to the average profile of the good prognosis group (middle panel). Solid line,

prognostic classifier with optimal accuracy; dashed line, with optimized sensitivity. Above the dashed line patients have a good prognosis signature, below the dashed line the prognosis signature is poor. The metastasis status for each patient is shown in the right panel: white indicates patients who developed distant metastases within 5 years after the primary diagnosis; black indicates patients who continued to be disease-free for at least 5 years. **c**, Same as for **b**, but the expression data matrix is for tumours of 19 additional breast cancer patients using the same 70 optimal prognostic marker genes. Thresholds in the classifier (solid and dashed line) are the same as **b**. (See Fig. 1 for colour scheme.)

evaluating its power for correct classification using the 'leave-one-out' method for cross-validation (see Supplementary Information). Classification was made on the basis of the correlations of the expression profile of the 'leave-one-out' sample with the mean expression levels of the remaining samples from the good and the poor prognosis patients, respectively. The accuracy improved until the optimal number of marker genes was reached (70 genes).

The expression pattern of the 70 genes in the 78 samples is shown in the colour plot of Fig. 2b (left panel), where tumours were ordered by rank according to their correlation coefficients with the average good prognosis profile (Fig. 2b, middle panel). The classifier predicted correctly the actual outcome of disease for 65 out of the 78 patients (83%), with respectively 5 poor prognosis and 8 good

prognosis patients assigned to the opposite category (Fig. 2b, threshold 'optimal accuracy', solid line). However, for the selection of patients eligible for adjuvant systemic therapy, a lower number of poor prognosis patients assigned to the good prognosis category should be attained. For this purpose, we set a threshold that resulted in misclassification of no more than 10% of the poor prognosis patients (3 patients out of 34 of the poor prognosis group). This optimized sensitivity threshold resulted in a total of 15 misclassifications: 3 poor prognosis tumours were classified as good prognosis, and 12 good prognosis tumours were classified as poor prognosis (Fig. 2b, dashed line). We classified tumours having a gene expression profile with a correlation coefficient above the 'optimized sensitivity' threshold (dashed line) as a good prognosis

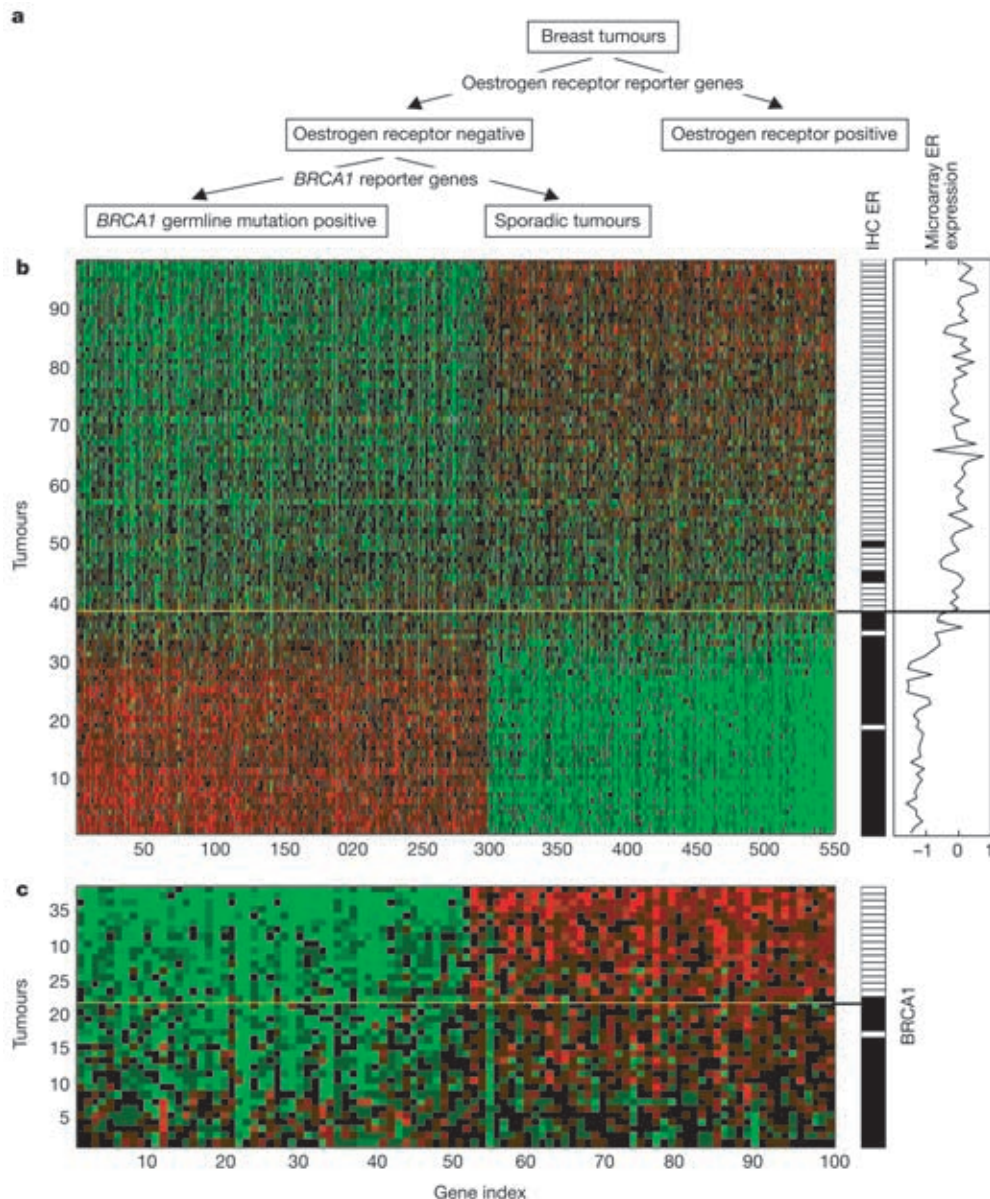


Figure 3 Supervised classification on ER and *BRCA1* signatures. **a**, Outline of a two-level classification system: 98 breast tumours are first classified into an ER-positive group and an ER-negative group, which is further divided into *BRCA1* mutation and sporadic tumours. **b**, Expression data matrix of the 98 sporadic tumours across 550 optimal ER reporter genes. The contrasting patterns discriminate between tumours with an ER-negative signature (below solid line) and an ER-positive signature (above solid line). The reporter genes were ordered on the basis of their level of contribution to the classifiers. Tumours are arranged according to the leave-one-out correlation coefficients to the

average signatures of the classifier. The ER status, as determined by IHC and microarray, are indicated in the two right panels. **c**, Expression data matrix of 38 ER-negative tumours defined by the ER classifier over the 100 optimal *BRCA1* reporter genes. The degree of the patterns divides the tumours in the ER-negative group into two subgroups: *BRCA1*-like and sporadic-like. Patients above the solid line are characterized by a *BRCA1* signature. The classification for each tumour was based on the leave-one-out procedure. The *BRCA1* germline mutation status is indicated in the right panel (white indicates mutation). (See Fig. 1 for colour scheme.)

signature, and below this threshold as a poor prognosis signature. Even small primary tumours without lymph node metastases can display the poor prognosis signature, indicating that they are already programmed for this metastatic phenotype.

The functional annotation for the genes provides insight into the underlying biological mechanism leading to rapid metastases. Genes involved in cell cycle, invasion and metastasis, angiogenesis, and signal transduction are significantly upregulated in the poor prognosis signature (for example cyclin E2, *MCM6*, metalloproteinases *MMP9* and *MP1*, *RAB6B*, *PK428*, *ESM1*, and the VEGF receptor *FLT1*; see Fig. 2b). If we evaluate all 231 prognostic reporter genes, more genes belonging to these functional categories become apparent (for example, *RAD21*, cyclin B2, *PCTAIRE*, *CDC25B*, *CENPF*, *VEGF*, *PGK1*, *MAD2*, *CKS2*, *BUB1*) (for a complete list, see Supplementary Information Table S2).

Many clinical studies have correlated alterations in expression of individual genes with breast cancer disease outcome, often with contradictory results. Examples include cyclin D1, ER- α , *UPA*, *PAI-1*, *HER2/neu* and *c-myc*^{19–22}. Surprisingly, none of these genes are present in our set of 70 marker genes. This could be due to the fact that here we determine gene expression at the level of transcription, whereas most previous studies measured protein levels. However, it is more likely that these genes in isolation have only limited predictive power, which highlights the need for an approach based on many genes.

To validate the prognosis classifier, an additional independent set of primary tumours from 19 young, lymph-node-negative breast cancer patients was selected. This group consisted of 7 patients who remained metastasis free for at least five years, and 12 patients who developed distant metastases within five years. The disease outcome was predicted by the 70-gene classifier and resulted in 2 out of 19 incorrect classifications using both the optimal accuracy threshold (Fig. 2c, solid line) and the optimized sensitivity threshold (Fig. 2c, dashed line). Thus, the classifier showed a comparable performance on the validation set of 19 independent sporadic tumours and confirmed the predictive power and robustness of prognosis classification using the 70 optimal marker genes (Fisher's exact test for association $P = 0.0018$).

The prediction of the classifier presented in Fig. 2b would indicate that women under 55 years of age who are diagnosed with lymph-node-negative breast cancer that has a poor prognosis signature have a 28-fold odds ratio (OR) (95% confidence interval, CI 7–107, $P = 1.0 \times 10^{-8}$) to develop a distant metastasis within 5 years compared with those that have the good prognosis signature (see Methods for odds ratio definition). This estimate, however, is based on the same series of patients that the classifier was derived from, and therefore this odds ratio represents an upper limit. A performance cross-validation procedure, in which the leave-one-out sample is not involved in selecting the prognosis reporter genes and the number of reporter genes is not optimized, results in an odds ratio of 15 for a short interval to metastases (95% CI 4–56, $P = 4.1 \times 10^{-6}$) (see Supplementary Information). This cross-validated predictive value of our classifier is superior to the currently available clinical and histopathological prognostic factors: high grade (odds ratio, OR = 6.4 (95% CI 2.1–19), $P = 0.0008$), tumour size greater than 2 cm (OR = 4.4 (95% CI 1.7–11), $P = 0.0028$), angioinvasion (OR = 4.2 (95% CI 1.5–12), $P = 0.01$), age ≤ 40 (OR = 3.7 (95% CI 1.3–11), $P = 0.02$), and ER negative (OR = 2.4 (95% CI 0.9–6.6), $P = 0.13$). Furthermore, the evaluation of the cross-validated classifier in a multivariate model that includes all classical prognostic factors indicates that it is an independent factor in predicting outcome of disease (logistic regression OR = 18 (3.3–94), P -value of likelihood ratio test 1.4×10^{-4}). Studying a large and unselected cohort of breast cancer patients is required to provide a more accurate estimate of the metastatic risk associated with the prognosis signature.

Unsupervised cluster analysis distinguishes between ER-positive

and ER-negative tumours (Fig. 1a). To investigate the expression patterns associated with the immunohistochemical staining of ER and to explore the differences between the sporadic and *BRCA1* tumours that fall into the ER-negative cluster (Fig. 1a), a supervised two-layer classification was performed (Fig. 3a). Figure 3b shows that 550 genes optimally report the dominant pattern associated with ER status, including genes such as keratin 18, *BCL2*, *ERBB3* and *ERBB4* (see Supplementary Information Table S3). The leave-one-out analysis shows that only two ER-positive and three ER-negative tumours (as determined by IHC) were classified in the opposite gene expression group (95% correct classification, Fig. 3b, middle panel). However, in all five discordant cases, the abundance of ER messenger RNA measured by the microarray agrees with the classification (Fig. 3b, right panel). An ER status reporter signature was also determined by others using a similar classification method⁸, and their ER signature gene set overlaps with ours (21 out of their 50 ER status reporter genes are present in our set of 550 ER reporters). Our observation in the unsupervised analysis that ER clustering has predictive power for prognosis is also valid for the ER supervised classification, although it does not reach the level of significance of the prognosis classifier (ER signature prediction for prognosis, OR = 3.7 (95% CI 1.3–11) $P = 0.02$; data not shown).

Figure 3c shows the leave-one-out classification of the 38 ER-negative tumours into sporadic cases and *BRCA1*-associated cases based on an optimal set of 100 genes. This set is enriched in lymphocyte-specific genes (see Supplementary Information Table S4). The classification into sporadic and *BRCA1* tumours was caused mainly by the differences in levels of gene expression (amplitude), in concordance with recent findings that *BRCA1* mediates ligand-independent transcriptional repression of the ER²³ (95% accuracy, 2/38 misclassified, Fig. 3c). The one sporadic tumour that was classified as a *BRCA1* tumour was shown to contain methylation of the *BRCA1* promoter, indicating an epigenetic modification of *BRCA1*²⁴ (data not shown). Notably, the discordant *BRCA1* tumour is from a patient where the germline mutation has only altered the last 29 amino acids of the *BRCA1* protein (*BRCA1* mutation 5,622del62), which abolishes transcriptional activation by *BRCA1*²⁵. One previous study defined a gene expression signature associated with *BRCA1* germline mutations using a panel of seven tumours²⁶; however, the study was unable to appreciate the overlap in signatures between the ER-negative and *BRCA1* tumours. Furthermore, the nine *BRCA1* status reporter genes²⁶ were not present in our set of 100 optimal reporter genes. The two-layer cluster analysis that we have used and the larger number of tumours we analysed may account for these differences.

Our results indicate that breast cancer prognosis can already be derived from the gene expression profile of the primary tumour. Recent consensus conferences on treatment of breast cancer in Europe and the USA (St. Gallen² and NIH consensus³) have developed guidelines for the eligibility of adjuvant chemotherapy based on histological and clinical characteristics. Following these

Table 1 Breast cancer patients eligible for adjuvant systemic therapy

Consensus	Patient group		
	Total patient group (<i>n</i> = 78)	Metastatic disease at 5 yr (<i>n</i> = 34)	Disease free at 5 yr (<i>n</i> = 44)
St Gallen	64/78 (82%)	33/34 (97%)	31/44 (70%)
NIH	72/78 (92%)	32/34 (94%)	40/44 (91%)
Prognosis profile*	43/78 (55%)	31/34 (91%)	12/44 (27%) (18/44 (41%)†)

The conventional consensus criteria are: tumour ≥ 2 cm, ER negative, grade 2–3, patient < 35 yr (either one of these criteria; St Gallen consensus); tumour > 1 cm (NIH consensus).

* Number of tumours having a poor prognosis signature using our microarray profile, defined by the optimized sensitivity threshold in the 70-gene classifier (see Fig. 2b).

† Number of tumours with a poor prognosis signature in the group of disease-free patients, when the cross-validated classifier is applied.

guidelines, up to 90% of lymph-node-negative young breast cancer patients are candidates for adjuvant systemic treatment. As 70–80% of these patients would not have developed distant metastases without adjuvant treatment, these patients may not benefit from the treatment, and may potentially suffer from the side effects. We applied the St Gallen and NIH consensus criteria on our patient group to compare the efficacy of the microarray classifier for the selection of patients for adjuvant systemic treatment. Table 1 shows that the prognosis classifier selects just as effectively those high-risk patients that would benefit from adjuvant therapy, but significantly reduces the number of patients that receive unnecessary treatment. Thus, the prognostic profile potentially provides a powerful tool to tailor adjuvant systemic treatment that could greatly reduce the cost of breast cancer treatment, both in terms of adverse side effects and health care expenditure. Furthermore, the signature that defines ER status can be used to decide on adjuvant hormonal therapy, and the signature that reveals *BRCA1* status may further improve the diagnosis of hereditary breast cancer. Finally, genes that are over-expressed in tumours with a poor prognosis profile are potential targets for the rational development of new cancer drugs. Identification of such targets may improve the efficiency of developing therapeutics for many tumour types. □

Methods

Breast tumour selection criteria

The criteria for the sporadic patients ($n = 97$) were: primary invasive breast carcinoma less than 5 cm (T1 or T2), no axillary metastases (N0), age at diagnosis less than 55 years, calendar year of diagnosis 1983–1996, no previous malignancies; all patients were treated by modified radical mastectomy ($n = 35$) or breast-conserving treatment ($n = 62$), including axillary lymph node dissection followed by radiotherapy. Five patients of the metastases group received adjuvant systemic therapy consisting of chemotherapy ($n = 3$) or hormonal therapy ($n = 2$), all other patients did not receive additional treatment. All patients were followed at least annually for a period of at least 5 years. The criteria for hereditary patients ($n = 20$) were: carriers of a germline mutation in *BRCA1* or *BRCA2*, and primary invasive breast carcinoma; no other selection criterion was applied. This study was approved by the Medical Ethical Committee of the Netherlands Cancer Institute. For complete patient data, see Table S1 in Supplementary Information.

Clinical parameters of breast tumours

Tumour material was snap-frozen in liquid nitrogen within 1 h after surgery. A haematoxylin and eosin stained section was prepared before and after cutting slides for RNA isolation for assessment of the percentage of tumour cells. Only samples with greater than 50% tumour cells were selected, mean 67% and median 70% for all groups studied. Formalin-fixed, paraffin-embedded tumour tissue was used to evaluate the following: tumour type (according to the World Health Organisation classification), histological grade (grade 1–3), and the presence of angioinvasive growth and extensive lymphocytic infiltrate. ER expression was determined by immunohistochemical staining (negative when less than 10% of the nuclei showed staining, all others ER positive).

RNA isolation

We used 30 sections of 30- μ m thickness for total RNA isolation. Total RNA was isolated with RNeasyLys, and finally dissolved in RNase-free H₂O. Twenty-five micrograms of total RNA was treated with DNase using the Qiagen RNase-free DNase kit and RNeasy spin columns. Total RNA treated with DNase was dissolved in RNase-free H₂O to a final concentration of 0.2 μ g μ l⁻¹.

cRNA labelling

cRNA was generated by *in vitro* transcription using T7 RNA polymerase on 5 μ g total RNA and labelled with Cy3 or Cy5 (CyDye, Amersham Pharmacia Biotech)¹³. Five micrograms of Cy-labelled cRNA from one breast cancer tumour was mixed with the same amount of reverse colour Cy-labelled product from a pool, which consisted of an equal amount of cRNA from each individual sporadic patient.

Expression profiling using microarray

Labelled cRNAs were fragmented to an average size of approximately 50–100 nucleotides by heating at 60 °C in the presence of 10 mM ZnCl₂, added to a hybridization buffer containing 1 M NaCl, 0.5% sodium sarcosine, 50 mM MES, pH 6.5, and formamide to a final concentration of 30%, final volume 3 ml at 40 °C. Hu25K microarrays represented the 24,479 biological oligonucleotides plus 1,281 control probes. Sequences for microarrays were selected from RefSeq (a collection of non-redundant mRNA sequences; <http://www.ncbi.nlm.nih.gov/LocusLink/refseq.html>) and from expressed sequence tag (EST) contigs (http://www.phrap.org/est_assembly/human/gene_number_methods.html). Each mRNA or EST contig was represented on the Hu25K microarray by a single 60-polymer

oligonucleotide chosen by the oligonucleotide probe design programme¹³. After hybridization, slides were washed and scanned using a confocal laser scanner (Agilent Technologies). Fluorescence intensities on scanned images were quantified, corrected for background noise and normalized¹³. Microarray data are available at <http://www.rii.com/publications/default.htm>.

Method of unsupervised two-dimensional clustering

In the two-dimensional cluster analysis, gene clustering and tumour clustering were performed independently using an agglomerative hierarchical clustering algorithm. For gene clustering, pairwise similarity metrics among genes are calculated on the basis of expression ratio measurements across all tumours. Similarly, for tumour clustering, pairwise similarity measures among tumours are calculated based on expression ratio measurements across all significant genes (for details see Supplementary Information).

Method of supervised classification

We developed a method for classifying breast tumours into prognostic or diagnostic categories based on gene expression profiles. This method includes the following three steps: (1) selection of discriminating candidate genes by their correlation with the category; (2) determination of the optimal set of reporter genes using a leave-one-out cross validation procedure; (3) prognostic or diagnostic prediction based on the gene expression of the optimal set of reporter genes (for details see Supplementary Information).

Statistical analysis

The odds ratio is the ratio of the odds in favour of developing distant metastases within 5 years for a patient in this study with a tumour characterized by the poor prognosis signature, to the odds in favour of developing metastases without this signature (2 $\times$ 2 table). *P*-values associated with odds ratios are calculated by Fisher's exact test. In the multivariate analysis a logistic model was applied with outcome of disease as the dependent variable, and the *P*-value for the relevant parameter is derived from the likelihood ratio test in the model (see Supplementary Information).

Received 24 August; accepted 22 November 2001.

- McGuire, W. L. Breast cancer prognostic factors: evaluation guidelines. *J. Natl Cancer Inst.* **83**, 154–155 (1991).
- Goldhirsch, A., Glick, J. H., Gelber, R. D. & Senn, H. J. Meeting highlights: international consensus panel on the treatment of primary breast cancer. *J. Natl Cancer Inst.* **90**, 1601–1608 (1998).
- Eifel, P. *et al.* National institutes of health consensus development conference statement: adjuvant therapy for breast cancer, November 1–3, 2000. *J. Natl Cancer Inst.* **93**, 979–989 (2001).
- Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. *Lancet* **352**, 930–942 (1998).
- Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* **351**, 1451–1467 (1998).
- Perou, C. M. *et al.* Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl Acad. Sci. USA* **96**, 9212–9217 (1999).
- Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- Gruvberger, S. *et al.* Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res.* **61**, 5979–5984 (2001).
- Martin, K. J. *et al.* Linking gene expression patterns to therapeutic groups in breast cancer. *Cancer Res.* **60**, 2232–2238 (2000).
- Zajchowski, D. A. *et al.* Identification of gene expression profiles that predict the aggressive behavior of breast cancer cells. *Cancer Res.* **61**, 5168–5178 (2001).
- Sortie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98**, 10869–10874 (2001).
- West, M. *et al.* Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc. Natl Acad. Sci. USA* **98**, 11462–11467 (2001).
- Hughes, T. R. *et al.* Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nature Biotechnol.* **19**, 342–347 (2001).
- Roberts, C. J. *et al.* Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* **287**, 873–880 (2000).
- Lakhani, S. R. *et al.* Multifactorial analysis of differences between sporadic breast cancers and cancers involving *BRCA1* and *BRCA2* mutations. *J. Natl Cancer Inst.* **90**, 1138–1145 (1998).
- Brenton, J. D., Aparicio, S. A. & Caldas, C. Molecular profiling of breast cancer: portraits but not physiognomy. *Breast Cancer Res.* **3**, 77–80 (2001).
- Khan, J. *et al.* Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Med.* **7**, 673–679 (2001).
- He, Y. D. & Friend, S. H. Microarrays—the 21st century divining rod? *Nature Med.* **7**, 658–659 (2001).
- Bieche, I. *et al.* Genetic alterations in breast cancer. *Genes Chromosomes Cancer* **14**, 227–251 (1995).
- Stee, P. S. & Zhou, Q. Cyclins and breast cancer. *Breast Cancer Res. Treat.* **52**, 17–28 (1998).
- Janicke, F. *et al.* Randomized adjuvant chemotherapy trial in high-risk, lymph node-negative breast cancer patients identified by urokinase-type plasminogen activator and plasminogen activator inhibitor type 1. *J. Natl Cancer Inst.* **93**, 913–920 (2001).
- van Diest, P. J. *et al.* Cyclin D1 expression in invasive breast cancer. Correlations and prognostic value. *Am. J. Pathol.* **150**, 705–711 (1997).
- Zheng, L., Annab, L. A., Afshari, C. A., Lee, W. H. & Boyer, T. G. *BRCA1* mediates ligand-independent transcriptional repression of the estrogen receptor. *Proc. Natl Acad. Sci. USA* **98**, 9587–9592 (2001).
- Esteller, M. *et al.* Promoter hypermethylation and *BRCA1* inactivation in sporadic breast and ovarian tumors. *J. Natl Cancer Inst.* **92**, 564–569 (2000).
- Chapman, M. S. & Verma, I. M. Transcriptional activation by *BRCA1*. *Nature* **382**, 678–679 (1996).
- Hedenfalk, I. *et al.* Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.* **344**, 539–548 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank D. Atsma and D. Majoor for assistance with the histological analyses and the preparation of tumour RNA; T. van der Velde, W. van Waardenburg and O. Dalesio for medical record data extraction; D. Slade, J. McDonald, J. Koch, T. Erkkila, M. Parrish and others at Rosetta's High Throughput Gene Expression Profiling Facility for microarray experiments; R. Stoughton, F. van Leeuwen, M. Rookus, P. Nederlof, F. Hogervorst and D. Voskuil for suggestions; and A. Berns, L. Hartwell, J. Radich and S. Rodenhuis for support and reading of the manuscript. This work was supported by a grant from the Center for Biomedical Genetics.

Competing interests statement

The authors declare competing financial interests: details accompany the paper on Nature's website (<http://www.nature.com>).

Correspondence and requests for materials should be addressed to S.H.F. (e-mail: stephen_friend@merck.com).

Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease

Laurent Monney*, Catherine A. Sabatos*, Jason L. Gaglia*, Akemi Ryu*, Hanspeter Waldner*, Tatyana Chernova†, Stephen Manning‡, Edward A. Greenfield*†, Anthony J. Coyle‡, Raymond A. Sobel§, Gordon J. Freeman† & Vijay K. Kuchroo*

* Center for Neurologic Diseases, Department of Neurology, Brigham and Women's Hospital and Harvard Medical School; and † Department of Adult Oncology, Dana-Farber Cancer Institute and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA

‡ Millennium Pharmaceuticals, 640 Memorial Drive, Cambridge, Massachusetts 02139, USA

§ VA Health Care System, Palo Alto and Department of Pathology, Stanford University School of Medicine, Stanford, California 95305, USA

Activation of naive CD4⁺ T-helper cells results in the development of at least two distinct effector populations, Th1 and Th2 cells¹⁻³. Th1 cells produce cytokines (interferon (IFN)- γ , interleukin (IL)-2, tumour-necrosis factor (TNF)- α and lymphotoxin) that are commonly associated with cell-mediated immune responses against intracellular pathogens, delayed-type hypersensitivity reactions⁴, and induction of organ-specific autoimmune diseases⁵. Th2 cells produce cytokines (IL-4, IL-10 and IL-13) that are crucial for control of extracellular helminthic infections and promote atopic and allergic diseases⁴. Although much is known about the functions of these two subsets of T-helper cells, there are few known surface molecules that distinguish between them⁶. We report here the identification and characterization of a transmembrane protein, Tim-3, which contains an immunoglobulin and a mucin-like domain and is expressed on differentiated Th1 cells. *In vivo* administration of antibody to Tim-3 enhances the clinical and pathological severity of experimental autoimmune encephalomyelitis (EAE), a Th1-dependent autoimmune disease, and increases the number and activation level of macrophages. Tim-3 may have an important role in the induction of autoimmune diseases by regulating macrophage activation and/or function.

In addition to their distinct roles in disease, Th1 and Th2 cells cross-regulate each other's expansion and functions. Thus, preferential induction of Th2 cells inhibits autoimmune diseases^{7,8}, and

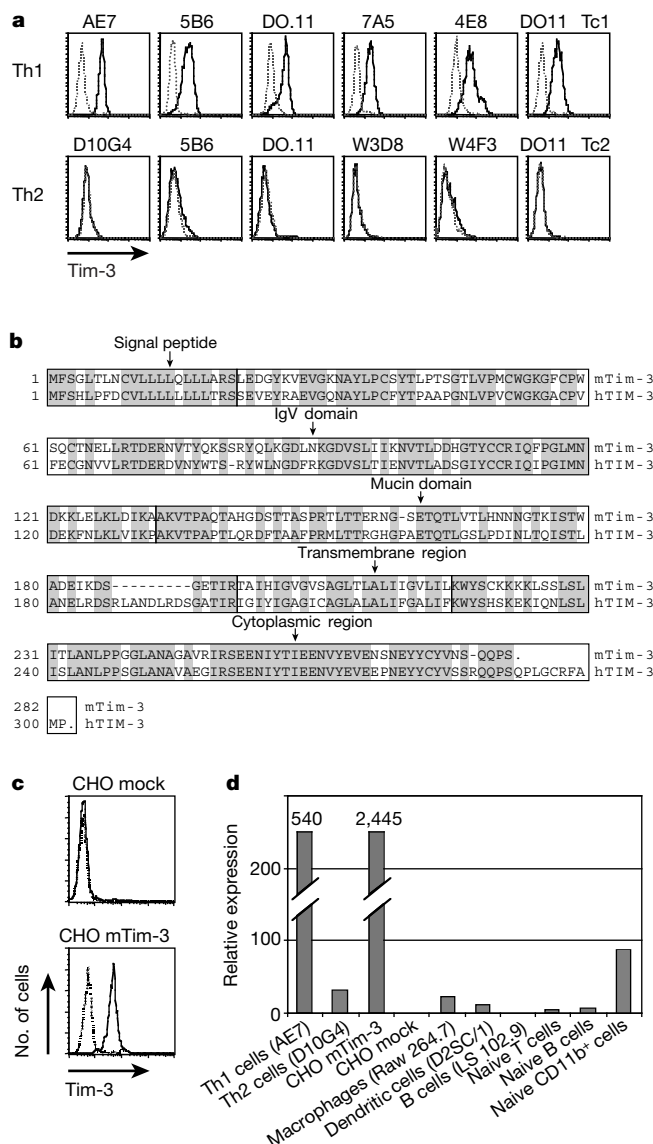


Figure 1 Cloning of a Th1-specific cell surface protein, Tim-3. **a**, Th1, Th2, Tc1 and Tc2 cells were stained with monoclonal antibody to Tim-3 (solid line) or rat IgG isotype control (dotted line). **b**, Deduced amino-acid sequence of murine and human Tim-3. Shading indicates regions of homology. IgV, variable region of immunoglobulin. **c**, CHO cells transfected using either Tim-3 cDNA (CHO mTim-3) or vector alone (CHO mock). Stable puromycin-resistant cells were stained with monoclonal antibody to Tim-3 (solid line) or rat IgG isotype control (dotted line). **d**, Total RNA from various cell lines and cells purified from SJL mice was isolated and transcribed to cDNA by reverse transcription, and cDNA was used for Taqman PCR. The figure shows expression of Tim-3 RNA relative to control GAPDH expression.

predominant induction of Th1 cells can regulate induction of asthma, atopy and allergies^{9,10}. Several groups have reported the association of chemokine and co-stimulatory receptors with Th1 (refs 11–14) and Th2 (refs 12, 13, 15–18) cells; however, the nature of the differences in expression of most of these molecules is quantitative.

To identify new Th1-specific cell surface proteins, we immunized Lewis and Lou/M rats with Th1 T-cell clones and lines, including the established Th1-specific clone AE7 and *in vitro* differentiated Th1 cell lines derived from 5B6 (ref. 19) and DO11.10 T-cell receptor (TCR) transgenic mice. A panel of approximately 20,000 monoclonal antibodies was generated and screened on Th1 and Th2 cells. Two of the monoclonal antibodies (8B.2C12 and 25F.1D6) that selectively stained Th1 cells were further characterized. These

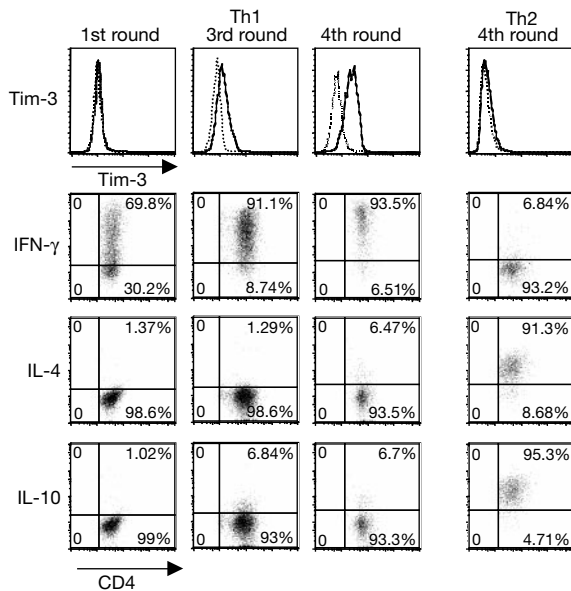


Figure 2 Tim-3 is detected on the surface of Th1 DO11.10 T cells after the third round of stimulation. Naive DO11.10 T cells were stimulated *in vitro* under Th1- and Th2-polarizing conditions. After each round of restimulation, Th1 and Th2 cells were stimulated with phorbol 12-myristate 13-acetate/ionomycin for the induction of cytokines and then stained with monoclonal antibodies to Tim-3 and CD4, and cytokine expression was detected by intracellular staining. Histograms represent Tim-3 expression on CD4⁺ cells (dotted line, isotype control; solid line, specific staining). Dot plots (logarithmic scale) represent cytokine expression in CD4⁺ cells.

monoclonal antibodies recognize a cell surface protein present on established CD4⁺ Th1 cells and CD8⁺ Tc1 (cytotoxic) cells, but absent on CD4⁺ Th2 cells and CD8⁺ Tc2 cells (Fig. 1a).

By gene-expression cloning²⁰, we identified a complementary DNA (Tim-3; T-cell immunoglobulin- and mucin-domain-containing molecule) encoding a type I membrane protein of 281 amino acids whose extracellular domain consists of an immunoglobulin variable-region-like domain followed by a mucin-like domain consisting of 31% serine and threonine residues. The cytoplasmic region contains six tyrosines, one of which is part of a tyrosine phosphorylation motif (RSEENIY). The extracellular domain contains four sites for N-linked and five sites for O-linked glycosylation. The human homologue of Tim-3 was identified through genomic database searches and polymerase chain reaction using reverse transcribed RNA (RT-PCR). It has a 63% amino-acid identity to murine Tim-3 overall, with 77% identity in the cytoplasmic domain, including conservation of the tyrosine phosphorylation motif (Fig. 1b). Database searches show that Tim-3 is related to kidney injury molecule-1 (Kim-1; also known as HAVcr-1), the receptor for human hepatitis A virus, and Tim-2;

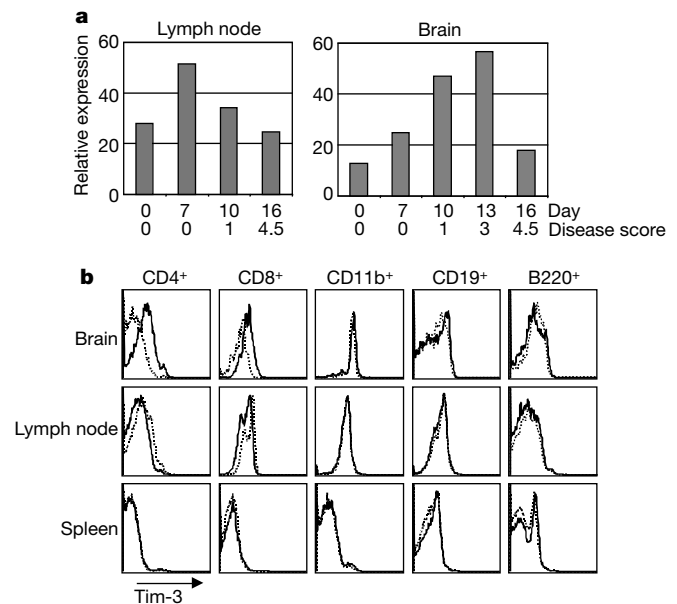


Figure 3 Tim-3 is expressed on CD4⁺ and CD8⁺ T cells during the course of EAE. SJL mice were immunized with PLP 139–151 peptide for the induction of EAE. Mice were killed at various time points after immunization, and spleen, brain and lymph nodes were removed and tested for the expression of Tim-3. **a**, Total RNA was isolated and transcribed to cDNA by reverse transcription and cDNA was used for Taqman PCR. The figure shows expression of Tim-3 relative to control GAPDH. **b**, On day 10, cells from brain, spleen and lymph nodes were stained with monoclonal antibodies to Tim-3, CD4, CD8, CD11b, CD19 and B220. Histograms represent Tim-3 expression on different cell populations (dotted line, isotype control; solid line, specific staining).

all of which have a similar immunoglobulin and mucin structure. All three family members (Kim-1, Tim-2 and Tim-3) are located on human chromosome 5q33.2 and mouse chromosome 11. A Tim-3 cDNA expression construct was used to stably transfect Chinese hamster ovary (CHO) cells. Flow cytometry and immunoprecipitation analysis of the stable transfectants confirmed that the cloned protein is Tim-3, and that it is expressed at the cell surface (Fig. 1c and data not shown).

Expression of this gene in various cell lines (Th1, Th2, dendritic, macrophage and B cells) and SJL T-, B- and CD11b⁺ cells was determined by quantitative RT-PCR. Tim-3 transcripts were present at the highest level only in Th1 cells, although *in vivo*-derived CD11b⁺ cells also showed low level of expression (Fig. 1d). Flow cytometric analysis using anti-Tim-3 antibodies confirmed that Tim-3 is not expressed on naive T cells, B cells, macrophages or dendritic cells (data not shown). These data suggest that the molecule recognized by these antibodies is expressed mainly on differentiated Th1 or Tc1 cells and not on other haematopoietic cell types, at least at the protein level. To examine the kinetics of Tim-3 protein expression during T-cell differentiation, naive DO11.10

Table 1 EAE in mice treated with anti-Tim-3 or control antibody

Treatment	Clinical EAE				Histological EAE			
	Incidence	Mortality	Day of onset	Mean maximal score	Incidence	Meningeal foci	Parenchymal foci	Total foci
Anti-Tim-3	28/29 (97%)	12/29* (41%)	11.5 $\pm$ 0.4†	3.6 $\pm$ 0.3‡	14/16 (88%)	53.2 $\pm$ 11.7	64.6 $\pm$ 14.9	117.8 $\pm$ 26.1
PBS	18/18 (100%)	2/18 (11%)	11.8 $\pm$ 0.6	2.7 $\pm$ 0.3	11/11 (100%)	33.5 $\pm$ 7.5	25.5 $\pm$ 9.4	59.1 $\pm$ 16.2
rlgG	23/26 (88%)	2/26 (8%)	15.4 $\pm$ 1.7	2.3 $\pm$ 0.3	13/15 (87%)	32.3 $\pm$ 8.3	38.5 $\pm$ 9.9	66.9 $\pm$ 17.8

* $P < 0.01$ when compared with the group treated with rat IgG (rlgG), and $P < 0.05$ when compared with the group treated with PBS.

† $P < 0.05$ when compared with the group treated with rlgG.

‡ $P < 0.01$ when compared with the group treated with rlgG, and $P < 0.05$ when compared with the group treated with PBS.

TCR transgenic T cells were activated *in vitro* under Th1- or Th2-polarizing conditions. Tim-3 is not expressed on Th2 cells but is detectable on Th1 cells after the third round of *in vitro* polarization (Fig. 2). Tim-3 is therefore expressed at a late stage of T-cell differentiation, suggesting that Tim-3 may not contribute to T-cell differentiation as such, but may have a function in the transport and/or effector functions of Th1 cells.

We next examined the expression of Tim-3 during the development of experimental autoimmune encephalomyelitis (EAE), a Th1-mediated autoimmune disease of the central nervous system (CNS). We immunized EAE-susceptible SJL mice with the encephalitogenic proteolipid protein (PLP) 139–151 peptide. Using quantitative RT-PCR, we detected that in the lymph node, Tim-3 messenger RNA was upregulated after immunization with expression peaking at 7 days after immunization, just before onset of EAE. Expression in the lymph nodes was then downregulated as the disease progressed. In the brain, expression of Tim-3 mRNA steadily increased and peaked at the onset of the disease (around day 10–13). Tim-3 mRNA expression was then downregulated to near basal levels at the maximal disease score (Fig. 3a). This expression pattern was confirmed at the protein level by flow cytometric analysis. Whereas Tim-3 is expressed on very few CD4⁺ T cells (less than 2%) in the periphery after immunization, Tim-3 is specifically expressed on most of the CD4⁺ and CD8⁺ T cells present in the CNS at the onset of clinical signs of disease (Fig. 3b). The number of Tim-3⁺ T cells decreased in the CNS as the disease progressed. These data suggest that Tim-3 is expressed *in vivo* on T cells and may have an important function in the initiation of EAE.

To study the function of Tim-3 *in vivo*, we tested the effect of anti-Tim-3 antibody administration on the development of EAE. SJL mice immunized with PLP 139–151 peptide were administered anti-Tim-3 antibody and monitored for the development of EAE. Mice treated with anti-Tim-3 developed a hyperacute and atypical EAE, with increased weight loss, malaise, ataxia and paralysis of the forelimbs, but without hindlimb paralysis and while sometimes retaining tail tone. Although disease onset was normal, progression was accelerated and resulted in significantly more severe clinical disease and increased mortality compared with the control group treated with control rat immunoglobulin- γ (IgG) (Table 1). Histologically, mice treated with anti-Tim-3 generally showed typical findings of acute EAE, but with increased numbers of inflammatory foci both in the meninges and parenchyma compared with the control group treated with rat IgG (Table 1). Animals treated with anti-Tim-3 that were killed immediately after the onset of clinical signs had a preponderance of neutrophils and mononuclear cells in the CNS inflammatory infiltrates with focal perivascular fibrin deposition, indicating vascular damage and a hyperacute inflammatory response (Fig. 4a, b). Animals that were killed at day 30 showed more extensive demyelinating lesions than animals treated with rat IgG (Fig. 4c–f). The demyelinating lesions in mice treated with anti-Tim-3 were filled with activated macrophages with detectable myelin fragments (Fig. 4d, inset). Activated macrophages are the cells primarily responsible for demyelination in EAE²¹; furthermore, depletion of activated macrophages leads to an inhibition of EAE²². Thus, it is possible that activated macrophages induced by anti-Tim-3 antibody treatment are responsible for the hyperacute disease phenotype and enhanced inflammation and demyelination.

To understand the function of Tim-3 *in vivo* and its observed role in EAE, we examined spleen and lymph node cells from SJL mice that were immunized with PLP 139–151 peptide and treated with anti-Tim-3 antibody or rat IgG control antibody. Whereas control mice treated with rat IgG showed a low basal (background) proliferative response (1,000–5,000 c.p.m.) and a dose-dependent increase in proliferation with the addition of specific antigen, the spleen cells from mice treated with anti-Tim-3 had 6–10 times the basal response in the absence of antigen, although the response to

specific antigen was very similar (Fig. 5a). Flow cytometric analysis of spleens from mice treated with anti-Tim-3 revealed a 2–3-fold increase in CD11b⁺ cell population (data not shown).

To confirm whether CD11b⁺ cells in mice treated with anti-Tim-3 were proliferating—and thereby contributing to the basal proliferative responses—5-bromodeoxyuridine (BrdU) was administered in the drinking water of immunized SJL mice treated with anti-Tim-3 or rat IgG control antibody. Flow cytometric analysis of the splenocytes from mice treated with anti-Tim-3 compared with rat IgG revealed an increase in the number of both BrdU positive and BrdU negative cells only in the CD11b⁺ population (Fig. 5b). Two-thirds of these CD11b⁺ cells incorporated BrdU and expressed higher levels of major histocompatibility complex (MHC) class II (Fig. 5b and data not shown), suggesting that these activated CD11b⁺ cells may have a principal role in the high basal response of mice treated with anti-Tim-3.

To determine the role of T cells and non-T cells, especially the CD11b⁺ cells, in the basal proliferative response, we purified T and non-T cells from spleen cells of mice treated with anti-Tim-3 or rat IgG2a. Although there was only a modest basal proliferative response from purified T cells, B cells or CD11b⁺ monocytes alone from mice treated with anti-Tim-3, the co-culturing of both B cells and CD11b⁺ cells with the anti-Tim-3-treated T cells recapitulated the high basal response (Fig. 5c). In comparison, the addition of B cells, monocytes or both B cells and monocytes to T cells from mice treated with rat IgG2a produced only an additive effect (Fig. 5c). When criss-cross proliferation experiments were

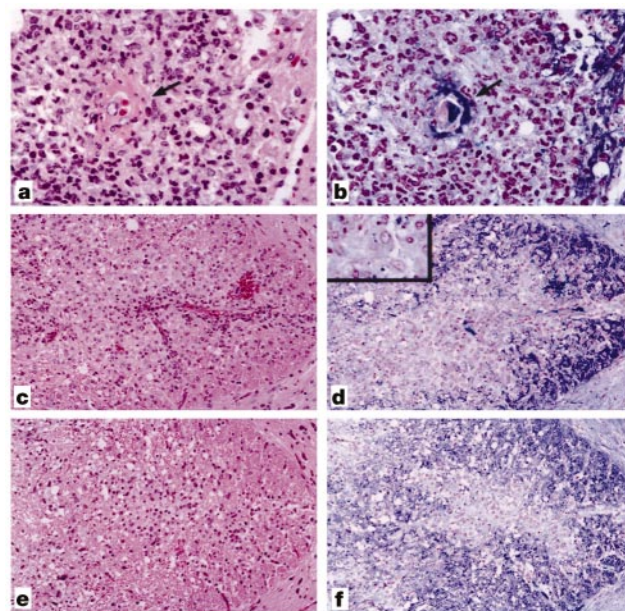


Figure 4 Treatment with anti-Tim-3 antibody enhances EAE. **a, b**, Inflammatory/demyelinating lesions in the spinal cord of a mouse treated with anti-Tim-3 on day 12 after immunization, at the peak of clinical disease. The infiltrate consists of a mixture of neutrophils and mononuclear cells. Perivascular fibrin deposition (arrow) indicates vasculitis and vascular injury. A portion of an intact peripheral nerve root (dark blue staining in **b**) is on the right side, whereas most of the CNS myelin in the field is lost. Magnification $\times 411$. **c, d**, Extensive demyelination associated with a perivascular mononuclear cell infiltrate in the posterior columns of the spinal cord of a mouse treated with anti-Tim-3 killed on day 30 after transfer. Inset, sheets of macrophages with phagocytosed myelin fragments (blue dots). Magnification $\times 137$; inset, $\times 411$. **e, f**, A similar infiltrate in the posterior columns of the spinal cord of a mouse treated with rat IgG killed on day 30 after immunization shows fewer macrophages and larger areas of intact (blue staining) myelin. Magnification $\times 137$. **a, c, e**, Staining with haematoxylin and eosin; **b, d, f**, Klüver–Barrera staining for myelin.

conducted, the addition of T cells treated with anti-Tim-3 to rat IgG2a-treated B cells plus CD11b⁺ cells increased the basal proliferative response; however, maximal basal proliferation required the presence of both anti-Tim-3-treated T cells and non-T cells. This basal proliferative response cannot be lowered by incubating the splenocytes with Fc-receptor (crystallizable fragment receptor, FcR)-blocking antibodies, suggesting that FcR-mediated crosslinking can be excluded as the mechanism for increasing basal response. The synergistic effect of the interaction between T cells and non-T cells is confirmed further, as addition of irradiated non-T cells to anti-Tim-3-treated T cells does not reconstitute the same high basal response. To determine whether a cognate interaction was required between T cells and non-T cells for the increase in the basal responses, we separated T cells and non-T cells with a permeable

membrane, which inhibits cell contact, and measured proliferative responses. As shown in Fig. 5c, separation of anti-Tim-3-treated T cells from anti-Tim-3-treated non-T cells resulted in a marked decrease in the proliferative responses, indicating that a cognate interaction between T and non-T cells is necessary.

To address directly whether T cells expressing Tim-3 regulate the expansion and activation of macrophages and other non-T cells, we adoptively transferred Tim-3⁺, Th1 5B6 transgenic T cells (with specificity for PLP 139–151 peptide) to naive SJL recipients, which were then immunized with PLP 139–151 peptide and treated with anti-Tim-3 or anti-ICOS antibody (as a T-cell antibody binding control). Flow cytometric analysis of spleen cells three days after transfer showed a marked increase in the number of activated CD11b⁺/F4/80⁺ macrophages in mice treated with anti-Tim-3,

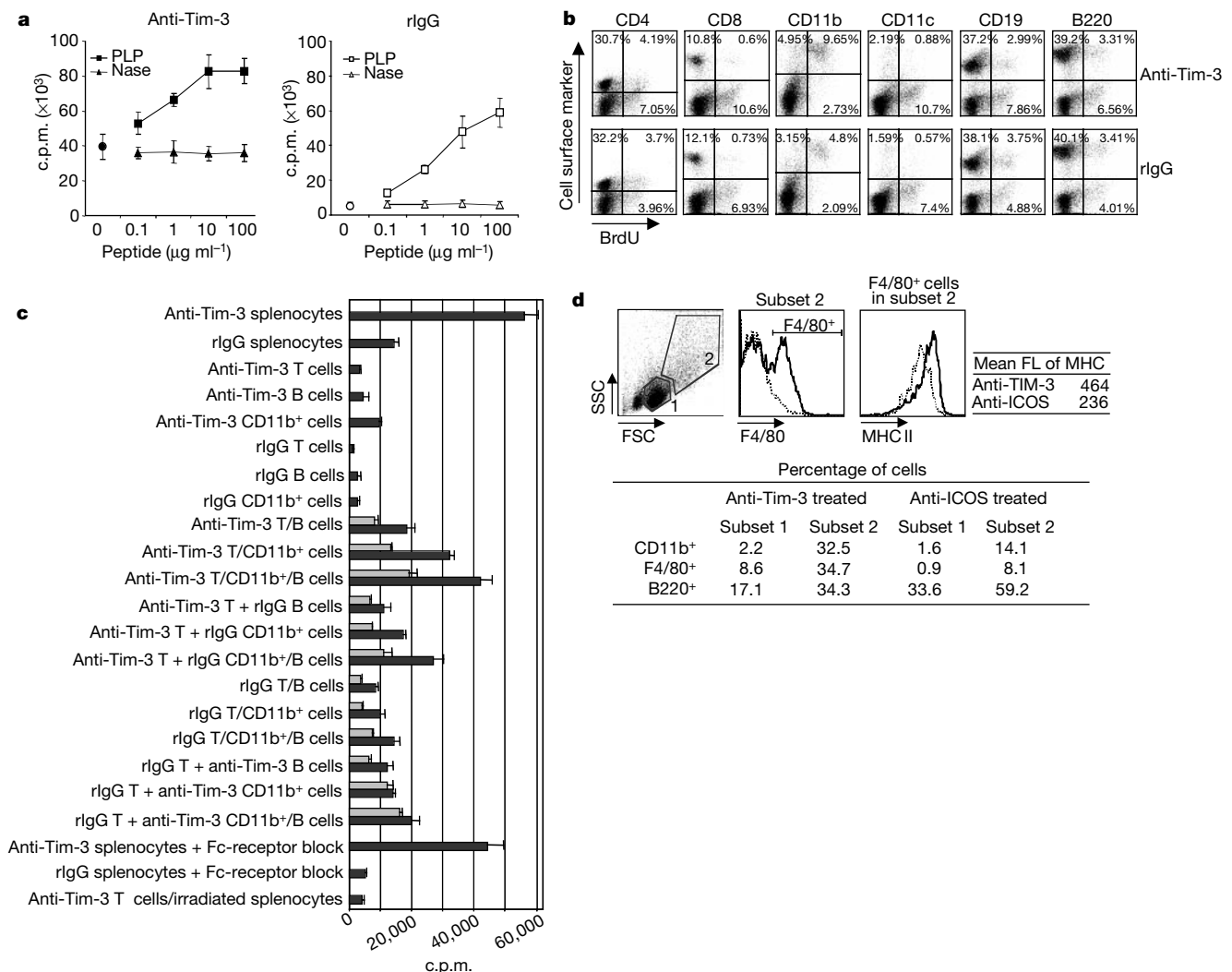


Figure 5 Treatment with anti-Tim-3 antibody induces macrophage activation. **a**, SJL mice immunized with PLP 139–151 peptide and given anti-Tim-3 or rat IgG (rIgG). Mice were killed on day 10, and spleens were removed. The cells were stimulated *in vitro* with increasing concentrations of PLP 139–151 peptide or neuraminidase 101–120 (Nase) peptide as control antigen. Proliferation was measured after 48 h by ³[H]-thymidine incorporation. **b**, SJL mice were given BrdU in their drinking water (0.8 mg ml⁻¹) after immunization with PLP 139–151 peptide and anti-Tim-3 or rat IgG antibody treatment. Mice were killed on day 10 and splenocytes were stained with monoclonal antibodies to CD4, CD8, CD11b, CD11c, CD19, B220 and BrdU. Dot plots (logarithmic scale) represent BrdU incorporation in the different cell populations. **c**, SJL mice immunized with PLP 139–151 peptide and given anti-Tim-3 or rat IgG2a treatment. Mice were killed on day 10. T, B and CD11b⁺ cells were purified from the spleen. Either whole splenocytes or T

cells (10^5) were incubated in the presence or absence of 2×10^5 B cells, 2×10^5 CD11b⁺ cells or irradiated splenocytes, and the proliferative response (black bars) was measured (³[H]-thymidine incorporation in triplicate wells). In a separate set of wells (grey bars), the purified T cells and non-T cells were separated by a permeable 0.2- μm Anapore membrane to eliminate cognate interaction. **d**, SJL mice were injected with 5×10^6 Tim-3-expressing, Th1 5B6 transgenic T cells specific for the PLP 139–151 peptide, then immunized with PLP 139–151 peptide. The recipients were also injected with anti-Tim-3 or anti-ICOS (as control) antibody on days 0 and 2 (subset 1 and 2, respectively). Spleen cells were removed on day 3 and the cells were stained with monoclonal antibodies to CD11b, F4/80, B220 and MHC class II. FSC, forward scatter; SSC, side scatter; FL, fluorescence.

but not in mice treated with anti-ICOS. There was a 2–3-fold increase in the number of CD11b⁺/F4/80⁺ cells present in subset 2 (Fig. 5d). These F4/80⁺ macrophages also expressed higher levels of MHC class II, indicating that they are activated to a greater extent.

Taken together, these data indicate that a cognate interaction between non-T cells and Tim-3-expressing Th1 cells is affected by anti-Tim-3 treatment, resulting in the expansion and activation of CD11b⁺/F4/80⁺ macrophages. Several possible mechanisms may explain our finding: (1) anti-Tim-3 may crosslink Tim-3 protein on the surface of differentiated Th1 cells *in vivo* and amplify the production of pro-inflammatory cytokines (IFN- γ and TNF), which in turn may induce activation of macrophages; (2) anti-Tim-3 antibody could enhance migration of differentiated Th1 cells into the brain where these cells may increase the cellular influx of macrophages from the circulation; (3) anti-Tim-3 could block a cognate interaction of Tim-3 with its potential inhibitory ligand on macrophages, thus leading to enhanced macrophage activation in the presence of pro-inflammatory cytokines produced by Th1 cells.

Furthermore, as Th1 and Th2 cells cross-regulate each other's functions, expression of Tim-3 on Th1 cells and subsequent transport of Tim-3-bearing Th1 cells to tissue sites might also have a role in the regulation of asthma and atopy. Recent studies have shown that Hba mice (BALB/c congenic mice carrying an interval on chromosome 11 from DBA/2) that are resistant to development of airway hyperreactivity show a strong linkage to the Tim-3/Kim locus and express polymorphic *Tim-3* and *Tim-1* genes²³. In summary, we have described a molecule that is selectively expressed on the surface of Th1 cells and have shown that engagement of this molecule during T-cell activation results in macrophage activation and increased severity of an autoimmune disease. These data, together with the results in asthma-resistant mice, suggest that the *Tim* gene family may have an important role in the regulation of autoimmunity and allergies. □

Methods

Generation of Th1-specific monoclonal antibodies

Lewis and Lou/M female rats (Harlan Sprague–Dawley) were immunized three times by a combination of subcutaneous injection with either $1\text{--}5 \times 10^7$ Th1-polarized T-cell clones and/or lines. The rats were boosted and four days later spleen cells were fused with myeloma cells (American Type Culture Collection number CRL8006) using polyethylene glycol 1450 and selection in HAT medium²⁴. The supernatants from the fusion plate wells were screened by flow cytometry on Th1 and Th2 cells. All hybridoma wells that gave a positive shift on Th1, but not Th2, T-cell clones were expanded and subcloned twice by limiting dilution.

Cloning procedure

A eukaryotic expression library was constructed using mRNA from the AE7 Th1 clone and the pAXF vector. We carried out library screening by expression cloning according to the method developed in ref. 20. Immunoselected individual plasmids were transfected into COS cells followed by indirect immunofluorescence staining with anti-Tim-3 antibody (8B.2C12). Positive clones were sequenced.

In vitro T-cell differentiation

CD62L high, CD44 low sorted naive DO11.10 T cells were stimulated *in vitro* with irradiated BALB/c spleen cells and OVA 323–339 peptide ($10\text{ }\mu\text{g ml}^{-1}$, Quality Controlled Biochemicals) for 7 days in the presence of mIL-12 (5 ng ml^{-1} , Pharmingen) and anti-mIL-4 ($10\text{ }\mu\text{g ml}^{-1}$) for Th1 differentiation and mIL-4 (10 ng ml^{-1} , R&D Systems) and anti-mIL-12 ($10\text{ }\mu\text{g ml}^{-1}$, Pharmingen) for Th2 differentiation. The cultures were supplemented with recombinant murine IL-2 (10 ng ml^{-1}) on days 2 and 4. All cells were collected on day 10 and used for further analysis.

Induction of EAE

Female SJL mice (4–8 weeks old) (Jackson Laboratory) were injected subcutaneously in each flank with 25–75 μg PLP 139–151 peptide (HSLGKWLGHDPKF) (Quality Controlled Biochemicals) in complete Freund's adjuvant (Difco) supplemented with 400 μg *Mycobacterium tuberculosis* (Difco). Each mouse was also injected intravenously with 100 ng pertussis toxin (List Biological Laboratories) in 0.1 ml PBS. Mice were injected intraperitoneally every other day beginning on day 0 and continuing for 10–16 days with either 100 μg anti-Tim-3 (endotoxin activity of 0.2–0.8 EU (endotoxin units) mg^{-1}) or 100 μg control rat IgG or PBS. Mice were examined daily for signs of EAE, which were

graded as follows: flaccid tail, 1; uneven gait and impaired righting reflex, 2; total hindlimb paralysis, 3; fore- and hindlimb paralysis, 4; and moribund, 5. At the peak of the disease or at the end of the experiment, brains and spinal cords were removed and fixed in 10% formalin and examined histopathologically for inflammation and demyelination.

Proliferation assays

Female SJL mice (4–8 weeks old) were injected subcutaneously in each flank with 50 μg PLP 139–151 peptide emulsified in complete Freund's adjuvant. Mice were injected intraperitoneally every other day with either 100 μg anti-Tim-3 or 100 μg control rat IgG or PBS. Mice were killed on day 10, and spleens and lymph nodes were removed. Cells were plated at 5×10^5 cells per well in round-bottomed 96-well plates with PLP 139–151 peptide or neuraminidase 101–120 peptide (EALVRQGLAKVAYVYKPNNT) (Nase; Quality Controlled Biochemicals) added at 0–100 $\mu\text{g ml}^{-1}$ for 48 h, and plates were pulsed with 1 μCi [³H]-thymidine per well for 12 h. We measured the incorporated radiolabelled thymidine using a Beta Plate scintillation counter (Wallac).

Quantitative Taqman RT-PCR

RNA was reverse-transcribed to cDNA. We performed real-time quantitative RT-PCR using the Taqman system (Applied Biosystems). The expression levels of Tim-3 and internal reference GAPDH were measured by multiplex PCR using probes labelled with 6-carboxyfluorescein (FAM) or VIC (Applied Biosystems), respectively. The primers and probes were designed in the 3' untranslated region of the 2.8-kilobase transcript using Primer Express v1.0 software (Applied Biosystems). The Tim-3 sequences are as follows: FAM probe, 5'-ACAGCTGCCTGCCCCAGTGCCC-3'; forward primer, 5'-GCCCGTGGACCTCAGTTTC-3'; reverse primer, 5'-TGGGAGCCAGCACAGATCA-3'. We purchased the GAPDH primer/probe set from Applied Biosystems. The simultaneous measurement of Tim-3–FAM and GAPDH–VIC permitted normalization of the amount of cDNA added per sample. We performed PCRs using the Taqman Universal PCR Master Mix and the ABI PRISM 7700 Sequence Detection System. A comparative threshold cycle (C_T) was used to determine gene expression relative to the no-tissue control (calibrator). Hence, steady-state mRNA levels were expressed as an *n*-fold difference relative to the calibrator. For each sample, the Tim-3 C_T value was normalized using the formula $\Delta C = C_{T\text{Tim-3}} - C_{T\text{GAPDH}}$. To determine relative expression levels, the following formula was used: $\Delta\Delta C_c = \Delta C_{(1)\text{sample}} - \Delta C_{(1)\text{calibrator}}$ and the value used to plot relative Tim-3 expression was calculated using the expression $2^{-\Delta\Delta C_c}$.

Received 13 August; accepted 28 November 2001.

- Mosmann, T. R., Cherwinski, H., Bond, M. W., Giedlin, M. A. & Coffman, R. L. Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *J. Immunol.* **136**, 2348–2357 (1986).
- Mosmann, T. R. & Sad, S. The expanding universe of T-cell subsets: Th1, Th2 and more. *Immunol. Today* **17**, 138–146 (1996).
- Abbas, A. K., Murphy, K. M. & Sher, A. Functional diversity of helper T lymphocytes. *Nature* **383**, 787–793 (1996).
- Sher, A. & Coffman, R. L. Regulation of immunity to parasites by T cells and T cell-derived cytokines. *Annu. Rev. Immunol.* **10**, 385–409 (1992).
- Liblau, R. S., Singer, S. M. & McDevitt, H. O. Th1 and Th2 CD4⁺ T cells in the pathogenesis of organ-specific autoimmune diseases. *Immunol. Today* **16**, 34–38 (1995).
- Syrbe, U., Sivek, J. & Hamann, A. Th1/Th2 subsets: distinct differences in homing and chemokine receptor expression? *Springer Semin. Immunopathol.* **21**, 263–285 (1999).
- Kuchroo, V. K. *et al.* B7-1 and B7-2 costimulatory molecules activate differentially the Th1/Th2 developmental pathways: application to autoimmune disease therapy. *Cell* **80**, 707–718 (1995).
- Nicholson, L. B., Greer, J. M., Sobel, R. A., Lees, M. B. & Kuchroo, V. K. An altered peptide ligand mediates immune deviation and prevents autoimmune encephalomyelitis. *Immunity* **3**, 397–405 (1995).
- Lack, G. *et al.* Nebulized but not parenteral IFN- γ decreases IgE production and normalizes airways function in a murine model of allergen sensitization. *J. Immunol.* **152**, 2546–2554 (1994).
- Hofstra, C. L. *et al.* Prevention of Th2-like cell responses by coadministration of IL-12 and IL-18 is associated with inhibition of antigen-induced airway hyperresponsiveness, eosinophilia, and serum IgE levels. *J. Immunol.* **161**, 5054–5060 (1998).
- Loetscher, P. *et al.* CCR5 is characteristic of Th1 lymphocytes. *Nature* **391**, 344–345 (1998).
- Bonocchi, R. *et al.* Differential expression of chemokine receptors and chemotactic responsiveness of type 1 T helper cells (Th1s) and Th2s. *J. Exp. Med.* **187**, 129–134 (1998).
- Sallusto, F., Lenig, D., Mackay, C. R. & Lanzavecchia, A. Flexible programs of chemokine receptor expression on human polarized T helper 1 and 2 lymphocytes. *J. Exp. Med.* **187**, 875–883 (1998).
- Venkataraman, C., Schaefer, G. & Schindler, U. Cutting edge: Chandra, a novel four-transmembrane domain protein differentially expressed in helper type 1 lymphocytes. *J. Immunol.* **165**, 632–636 (2000).
- Jourdan, P. *et al.* IL-4 induces functional cell-surface expression of CXCR4 on human T cells. *J. Immunol.* **160**, 4153–4157 (1998).
- Zingoni, A. *et al.* The chemokine receptor CCR8 is preferentially expressed in Th2 but not Th1 cells. *J. Immunol.* **161**, 547–551 (1998).
- McAdam, A. J. *et al.* Mouse inducible costimulatory molecule (ICOS) expression is enhanced by CD28 costimulation and regulates differentiation of CD4⁺ T cells. *J. Immunol.* **165**, 5035–5040 (2000).
- Lohning, M. *et al.* T1/ST2 is preferentially expressed on murine Th2 cells, independent of interleukin 4, interleukin 5, and interleukin 10, and important for Th2 effector function. *Proc. Natl Acad. Sci. USA* **95**, 6930–6935 (1998).
- Waldner, H., Whitters, M. J., Sobel, R. A., Collins, M. & Kuchroo, V. K. Fulminant spontaneous autoimmunity of the central nervous system in mice transgenic for the myelin proteolipid protein-specific T cell receptor. *Proc. Natl Acad. Sci. USA* **97**, 3412–3417 (2000).
- Seed, B. & Aruffo, A. Molecular cloning of the CD2 antigen, the T-cell erythrocyte receptor, by a rapid immunoselection procedure. *Proc. Natl Acad. Sci. USA* **84**, 3365–3369 (1987).

21. Gordon, E. J., Myers, K. J., Dougherty, J. P., Rosen, H. & Ron, Y. Both anti-CD11a (LFA-1) and anti-CD11b (MAC-1) therapy delay the onset and diminish the severity of experimental autoimmune encephalomyelitis. *J. Neuroimmunol.* **62**, 153–160 (1995).
22. Tran, E. H., Hoekstra, K., van Rooijen, N., Dijkstra, C. D. & Owens, T. Immune invasion of the central nervous system parenchyma and experimental allergic encephalomyelitis, but not leukocyte extravasation from blood, are prevented in macrophage-depleted mice. *J. Immunol.* **161**, 3767–3775 (1998).
23. McIntire, J. J. *et al.* Identification of *Tapr* (an airway hyperreactivity regulatory locus) and the linked *Tim* gene family. *Nature Immunol.* **2**, 1019–1116 (2001).
24. Kohler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497 (1975).

Acknowledgements

We thank L. Nicholson, N. Ruggli, L. Glimcher and H. Cantor for comments on this manuscript; R. Nazareno and E. Howard for technical help and D. Schlesinger for discussions. This work was supported by grants from the National Multiple Sclerosis Society, New York, the Swiss National Science Foundation, and the National Institutes of Health (M.D.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.K.K. (e-mail: vkuchroo@rics.bwh.harvard.edu). The Tim-3 mRNA sequences have been deposited in GenBank under accession numbers AF450241–AF450243.

Phospholipase C γ 1 is a physiological guanine nucleotide exchange factor for the nuclear GTPase PIKE

Keqiang Ye*, Bahman Aghdasi*, Hongbo R. Luo*, John L. Moriarity*, Frederick Y. Wu*, Jenny J. Hong*, K. Joseph Hurt*, Sun Sik Bae†, Pann-Ghill Suh† & Solomon H. Snyder*

* Johns Hopkins University School of Medicine, Departments of Neuroscience, Pharmacology and Molecular Sciences and Psychiatry, 725 N. Wolfe Street, Baltimore, Maryland 21205, USA

† Department of Life Science, Division of Molecular Life Science, Pohang University of Science and Technology, San 31 Hyojadong, Pohang 790-784, Korea

Phospholipase C γ 1 (PLC- γ 1) hydrolyses phosphatidylinositol-4,5-bisphosphate to the second messengers inositol-1,4,5-trisphosphate and diacylglycerol. PLC- γ 1 also has mitogenic activity upon growth-factor-dependent tyrosine phosphorylation^{1,2}; however, this activity is not dependent on the phospholipase activity of PLC- γ 1, but requires an SH3 domain^{3,4}. Here, we demonstrate that PLC- γ 1 acts as a guanine nucleotide exchange factor (GEF) for PIKE (phosphatidylinositol-3-OH kinase (PI(3)K) enhancer). PIKE is a nuclear GTPase that activates nuclear PI(3)K activity, and mediates the physiological activation by nerve growth factor (NGF) of nuclear PI(3)K activity⁵. This enzymatic activity accounts for the mitogenic properties of PLC- γ 1.

PIKE possesses proline-rich domains, which typically bind to SH3 domains of other proteins^{6–8}. To identify PIKE-interacting proteins, we conducted pull-down experiments exploring potential interactions between PIKE and SH3 domains of proteins such as the p85 subunit of PI(3)K, Grb2, Fyn and PLC- γ (Fig. 1; see also Supplementary Information Fig. 1). The most robust interaction occurs with the SH3 domain of PLC- γ , although substantial binding is also evident for Fyn and the amino-terminal SH3 domain of Grb2; however, we observed no binding with p85 α or - β . The carboxy-terminal domain of protein 4.1N, which binds PIKE through a non-SH3 domain⁵, also binds PIKE in these experiments.

To test whether these different SH3 domains have any effect on

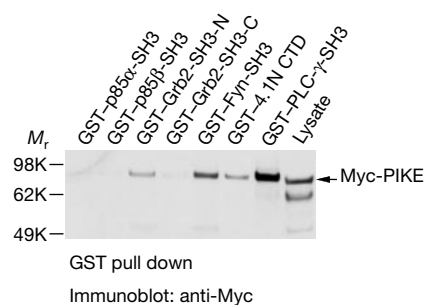


Figure 1 Binding of Myc-PIKE to different GST-SH3 domain fusion proteins. GST-SH3 fusion proteins were coupled to glutathione Sepharose beads and incubated with HEK 293 cell lysates transfected with Myc-PIKE. The beads were washed extensively and bound proteins eluted by boiling in Laemmli sample buffer. The samples were resolved by SDS-polyacrylamide gel electrophoresis (PAGE) on a 4–12% gel, and visualized by immunoblotting with anti-Myc antibody. Similar levels of various GST fusion proteins were used (see Supplementary Information Fig. 1). The relative molecular mass (M_r) is indicated along the left. CTD, C-terminal domain.

activation of the GTPase PIKE, we monitored the influence of their glutathione *S*-transferase (GST)-SH3 fusion proteins on the binding by PIKE of [³⁵S]GTP- γ S and [³H]GDP (Fig. 2a, b). PLC- γ -SH3 (residues 790–850) markedly augments the binding by PIKE of [³⁵S]GTP- γ S and accelerates the dissociation of [³H]GDP. By contrast, Fyn and Grb2 are inactive. A mutation at position 842 of proline to leucine in PLC- γ -SH3, which abolishes association with its target protein⁹, also diminishes its interaction with PIKE (Supplementary Information Fig. 2a). Consequently, GST-PLC- γ 1-SH3 (P842L) fails to stimulate binding by PIKE of [³⁵S]GTP- γ S or the dissociation of [³H]GDP (Fig. 2a, b). Selectivity for the SH3 domain of PLC- γ is evident from the enhancement of [³H]GDP dissociation from PIKE elicited by GST fusion proteins containing only the SH3

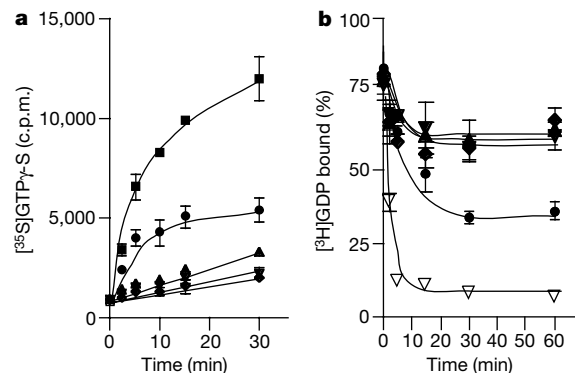


Figure 2 The SH3 domain of PLC- γ functions as a nucleotide exchange factor for PIKE. **a**, GST-PLC- γ -SH3 stimulates GTP binding to PIKE. PIKE (50 nM) was incubated with 50 nM GST-PLC- γ -SH3 (squares), GST-PLC- γ -SH3 (P842L) (circles), GST-Fyn-SH3 (diamonds), GST-Grb2-SH3-N (triangles) protein or control buffer (inverted triangles) for 30 min at 37 °C. [³⁵S]GTP- γ S (2 μ Ci ml⁻¹) (3 μ M) was added to start the assay. At the indicated time points, [³⁵S]GTP- γ S bound to PIKE was measured by a filter-binding assay. c.p.m., Counts per minute. **b**, Effects of purified GST-PLC- γ -SH3 fusion protein on the kinetics of GDP dissociation from PIKE. His-PIKE (50 nM) was preloaded with 0.5 μ M [³H]GDP for 30 min at 30 °C. Purified GST-PLC- γ -SH3 (open, inverted triangles), GST-PLC- γ -SH3 (P842L) (filled circles), GST-Fyn-SH3 (filled diamonds), GST-Grb2-SH3-N (filled triangles) proteins (50 nM each) or control buffer (filled, inverted triangles) were added together with 0.5 mM unlabelled GTP at the start of the assay. At the time intervals indicated, the PIKE-bound radioactivity was measured by a filter-binding assay. Stimulation of GDP dissociation is expressed as the percentage of [³H]GDP bound to PIKE before the addition of unlabelled GTP (see Supplementary Information). Standard error is indicated ($n = 3$).

domain or the SH3 and the two SH2 domains, but not by a construct containing only two SH2 domains (Supplementary Information Fig. 2b). GTP typically dissociates GEF proteins from their targets^{10,11}. GTP γ S dissociates the binding of PIKE to a construct expressing the two SH2 domains and the SH3 domain of PLC- γ (PLC-SH223) (Supplementary Information Fig. 2c). We also observed that GTP γ S showed a similar effect on the binding of PIKE to GST-PLC- γ 1-SH3 (data not shown).

SH3 domains bind typically to proline-rich domains of interacting proteins. To evaluate the portion of PIKE that binds to PLC- γ -SH3, we examined binding by various truncations of PIKE to PLC- γ using co-immunoprecipitation (Fig. 3a–c). Both amino acids 262–384 and the third proline-rich domain (353–362) bind robustly to full-length PLC- γ . In contrast, amino acids 262–384 weakly associate with overexpressed PLC- γ - Δ SH223, but the proline-rich domain itself does not bind to PLC- γ - Δ SH223 (Supplementary Information Fig. 3). To test whether the proline-rich domain of PIKE directly binds the SH3 domain of PLC- γ , we purified His-PIKE (262–384), which contains the proline-rich domain, and GST-PLC- γ -SH3. PLC- γ -SH3 strongly associates with His-PIKE (262–384) but not to a His-NuMA (residues 1440–1913) fragment, which was used as a control (Supplementary Information Fig. 3).

PIKE is localized to the nucleus⁵, therefore the pool of PLC- γ 1 that interacts with it should also be nuclear. Previous studies indicate that a portion of cellular PLC- γ 1 is nuclear^{12–17}. In subcellular fractionation experiments with PC12 cells, NGF exposure elicits a modest increase of endogenous PLC- γ in the nuclear fraction (Supplementary Information Fig. 4a). This is not associated with decreased cytoplasmic levels, presumably because of the extremely high cytoplasmic concentration of PLC- γ . To further confirm the endogenous PLC- γ nuclear translocation, we conducted immunofluorescent staining of PLC- γ . After 5–10 min of NGF treatment, PLC- γ translocates into the nuclei in both the native PC12 cells and PC12 cells stably transfected with wild-type PLC- γ (PLC- γ -WT) (Supplementary Information Fig. 4b). Western blotting reveals higher levels of PLC- γ in the nuclear fraction of PLC- γ -WT stably transfected PC12 cells than in native cells without

NGF treatment (data not shown).

NGF activates the GTP binding of PIKE in PC12 cells with activity peaking at 30 min and declining at 4 h (Fig. 4a). Activation of GTPases is linked to the dissociation of their GEFs^{10,11}. After NGF treatment of PC12 cells, PLC- γ forms a strong complex with PIKE at 5 min, subsequently dissociates from PIKE at 30 min and at 1 h, while binding returns at 4 h when PIKE activation diminishes (Fig. 4b). The binding of PLC- γ and PIKE is dependent on the SH3 domain of PLC- γ being diminished by exogenous GST-PLC- γ -SH3. This finding indicates a physiological regulation by NGF of interactions of PIKE with PLC- γ .

To explore further how NGF physiologically regulates the GEF activity of PLC- γ for PIKE, we used PC12 cells stably transfected with an inducible form of wild-type PLC- γ , PLC- γ with deletion of one of the SH2 domains, PLC- γ with deletion of the SH3 domain, or PLC- γ with a proline-to-leucine mutation at position 842 in the SH3 domain (Fig. 5). In control cells, PIKE is activated almost fourfold at 30 min after NGF treatment, confirming our earlier findings⁵. Cells transfected with wild-type PLC- γ display a threefold augmentation of PIKE activation even in the absence of NGF treatment, consistent with the GEF activity of PLC- γ . NGF treatment of these cells leads to a further enhancement of PIKE activation. A lipase-inactive mutant displays similar effects as the wild type both in the absence and presence of NGF, consistent with earlier findings^{3,4} that lipase activity is not necessary for the mitogenic activity of PLC. The catalytic activity of PLC- γ is not required for PIKE activation, which is unaffected by the PLC inhibitor U73122 (data not shown). Deletion of the SH2 domain of PLC- γ has no influence on PIKE activation in the presence or absence of NGF. By contrast, deletion of the SH3 domain of PLC- γ abolishes NGF activation of PIKE and greatly reduces PIKE activation in the absence of NGF (Fig. 5b). Notably, the single point mutant (P842L) displays effects that are similar to the deletion of the SH3 domain. These findings establish that activation by NGF of PIKE is dependent on PLC- γ -SH3.

PIKE physiologically augments nuclear PI(3)K activity⁵. Accordingly, if PLC- γ -SH3 physiologically regulates PIKE activation, it should also regulate nuclear PI(3)K activity. We monitored nuclear PI(3)K activity in the same PC12 cells with various PLC constructs as used for the PIKE activation studies (Fig. 5c). As observed for PIKE activation, PI(3)K activity in control cells is markedly augmented by NGF treatment. Overexpression of PLC- γ -WT increases PI(3)K activity even in the absence of NGF, while substantial further increases occur with NGF treatment. The lipase-inactive mutant displays similar effects as the wild type. Deletion of one of the SH2 domains has no effect, whereas deletion of the SH3 domain of PLC- γ

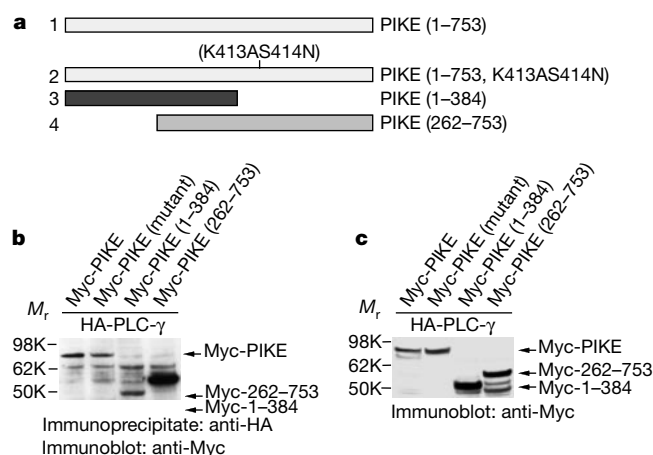


Figure 3 PIKE associates with PLC- γ 1. **a**, Diagram of PIKE constructs used in the co-immunoprecipitation experiments. **b**, The third proline-rich domain of PIKE (residues 353–362) is required for PIKE to associate with PLC- γ 1. HA-PLC- γ 1 and Myc-PIKE constructs were cotransfected into HEK 293 cells. PLC- γ 1 was immunoprecipitated with anti-HA antibody, and bound proteins were visualized by western blot with anti-Myc antibody. Wild-type PIKE, lysine-to-alanine and serine-to-asparagine-mutated (mutant) PIKE, PIKE (1–384) and PIKE (262–753) all bind to PLC- γ 1, which suggests that the amino-acid sequence 262–384, containing the third proline-rich domain, is involved in PIKE binding to PLC- γ 1. **c**, For experiments shown in **b**, similar levels of all Myc-PIKE constructs were expressed. The third proline-rich domain of PIKE binds to the SH3 domain of PLC- γ 1 (see Supplementary Information).

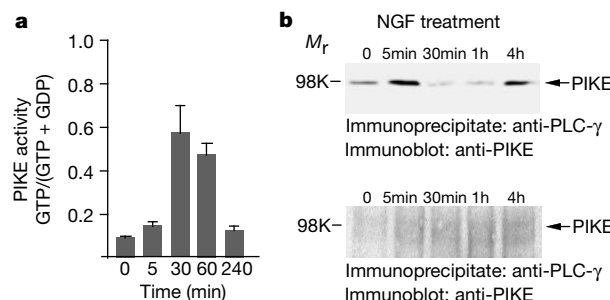


Figure 4 PLC- γ 1 associates with PIKE in PC12 cells in a GDP-dependent way. **a**, PIKE is activated by NGF at about 30 min treatment. **b**, The SH3 domain of PLC- γ 1 binds to PIKE in PC12 cells. PC12 cells were treated with 50 ng ml⁻¹ NGF and, at the indicated time points, cells were lysed and immunoprecipitated with mouse monoclonal anti-PLC- γ 1 antibody raised against the C-terminal portion. Co-precipitated PIKE was visualized by western blotting. In the presence of 10 μ g GST-PLC- γ -SH3 in the cell lysate, the interaction between PLC- γ and PIKE is diminished. Standard error is indicated ($n = 3$).

abolishes NGF activation of PI(3)K and profoundly reduces its activation in the absence of NGF. Consistent with its defects in activating PIKE, the PLC- γ mutant (P842L) fails to increase nuclear PI(3)K activity.

PLC- γ is physiologically activated after tyrosine phosphorylation by a variety of growth factor receptors for which PLC- γ is a crucial mediator of mitogenic actions¹⁸. Despite the importance of PLC- γ for normal and malignant cell growth, the molecular mechanisms underlying its mitogenic activity have been elusive. Phospholipase catalytic activity does not appear to be required, whereas the SH3 domain has been reported to be critical for mitogenesis in fibroblasts^{3,19}. To examine the role of the SH3 domain for mitogenesis in PC12 cells, we monitored mitogenesis by immunofluorescent

staining of 5-bromodeoxyuridine (BrdU) incorporation (Fig. 5d, e). Induction of PLC- γ markedly increases BrdU incorporation. This mitogenic activity is greatly reduced in cells lacking the SH3 domain or with the (P842L) mutation. A modest level of mitogenesis in uninduced cells is also diminished by deleting the SH3 domain of PLC- γ or by causing the (P842L) mutation. Our data show that the SH3 domain of PLC- γ is critical for mitogenesis, probably through the activation of the nuclear PI(3)K.

Our finding that PLC- γ displays physiological GEF activity for PIKE provides a mechanism that can explain the mitogenic activity of PLC- γ -SH3. Conceivably, such mitogenic activity is associated with the activation by PIKE of nuclear PI(3)K. If the mitogenic actions of PLC- γ stem from its PIKE-GEF activity, then both effects

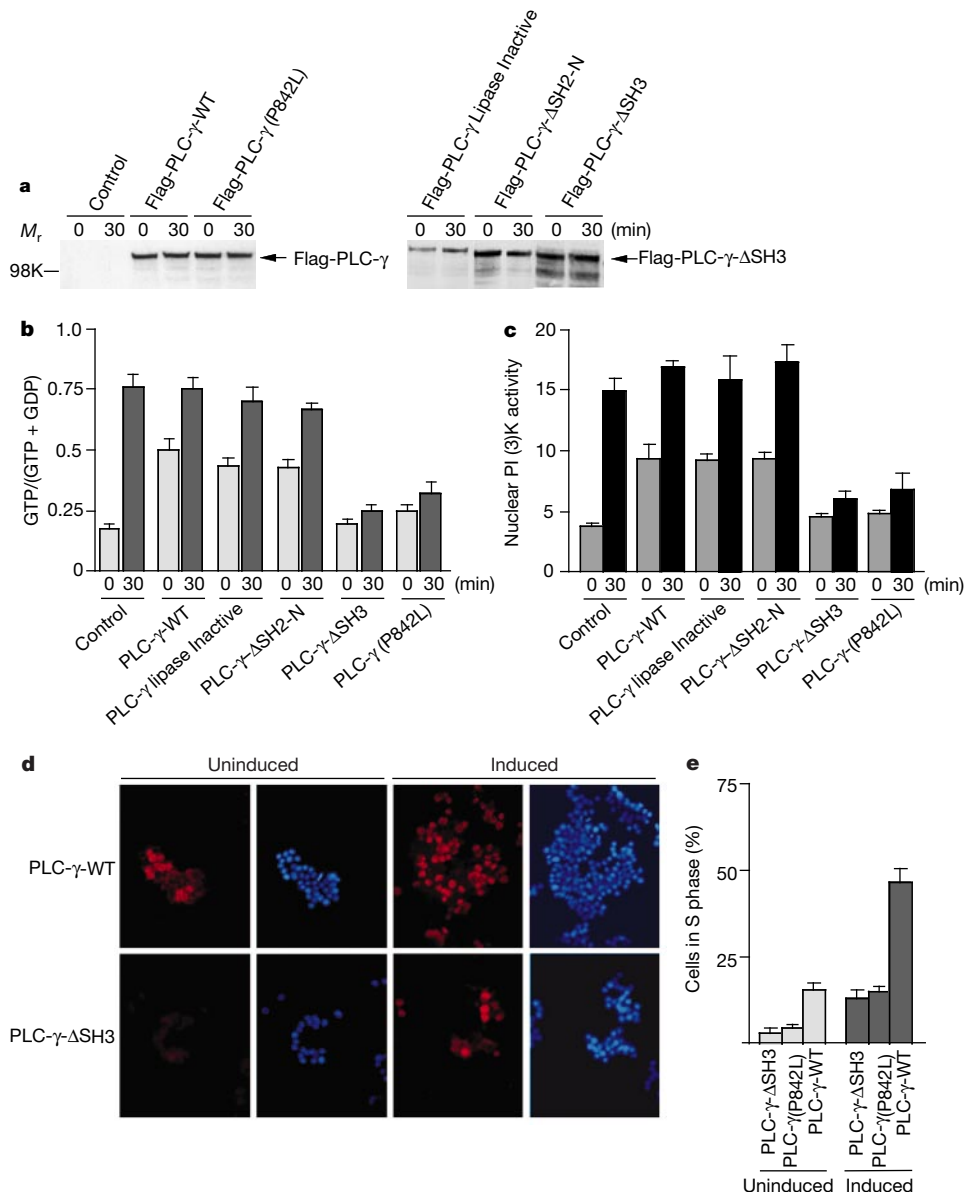


Figure 5 The SH3 domain of PLC- γ is required for the activation of PIKE and nuclear PI(3)K. **a**, Induction of Flag-tagged PLC- γ 1 in stably transfected PC12 cells. PC12 cells stably transfected with varying PLC- γ 1 constructs were induced by removing tetracycline from the culture medium for 2 d, then treated with 50 ng ml⁻¹ NGF for 30 min. Samples were resolved on 4–12% SDS–PAGE and analysed by immunoblotting with anti-Flag antibody. **b**, The SH3 domain of PLC- γ 1 is required for enhancement of GTP loading on PIKE. The stably transfected PC12 cells were induced to express different PLC- γ 1 constructs by removing tetracycline from the culture medium for 2 d, then the cells were metabolically labelled in phosphate-free DMEM with 0.5 mCi ml⁻¹ [³²P]H₃PO₄ for 4 h at

37 °C, followed by 30 min of NGF treatment. PIKE was immunoprecipitated with polyclonal anti-PIKE antibody, and bound guanine nucleotides were eluted and separated on TLC plates. The percentage of GTP was calculated as GTP/(GTP + GDP). **c**, The SH3 domain of PLC- γ 1 is required for activation of nuclear PI(3)K. The induced PC12 cells were treated with 50 ng ml⁻¹ NGF for 30 min, nuclei were isolated and lysed. p110 was immunoprecipitated from the lysate with rabbit polyclonal anti-p110 antibody. PI(3)K assays were performed on the immunoprecipitate. **d**, **e**, BrdU incorporation (red) into the nucleus of PC12 cells, which were stably transfected with inducible wild-type PLC- γ , PLC- γ - Δ SH3 and PLC- γ -(P842L). Blue shows DAPI staining. Standard error is indicated ($n = 3$).

should display the same structural requirements, that is, selective dependence on the SH3 group of PLC- γ . Evidence implicating PLC- γ -SH3 in PC12 cell growth comes from experiments with the same cells used in this study¹. Cells containing constructs deleting the SH3 domain fail to grow in the presence of serum, whereas cells with constructs containing only the SH3 and the two SH2 domains grow as well as cells transfected with wild type PLC- γ . Both SH223 domains and wild-type PLC- γ inhibit NGF-induced neurite outgrowth in transfected PC12 cells¹, establishing a requirement of PLC- γ -SH3 for neurotrophic actions of NGF. In another cell line, transfection with PLC- γ -SH3 alone promoted cell growth in NIH 3T3 cells and a catalytically inactive mutant of PLC- γ elicited a full mitogenic response^{3,4}.

In response to upstream signals, GEFs stimulate GDP release from their bound GTPases. The catalytic domains of the different classes of GEFs share no sequence homology and are structurally unrelated, although they have very similar substrates (small GTPases) and functions (the dissociation of GDP)^{20,21}. SH3 domains are generally regarded as binding motifs; however, PLC- γ -SH3 can display GTPase-activating actions on dynamin⁷. Moreover, PLC- γ -SH3 has been shown to augment GTP loading activity of the GTPase Ras by binding to its guanine nucleotide exchange protein Sos1 (refs 22–24). Our finding that PLC- γ -SH3 is a robust GEF for PIKE provides further evidence indicating that this domain can display enzymatic activity. □

Methods

Cells and reagents

PC12 cells were maintained in medium A (DMEM with 10% fetal bovine serum (FBS), 5% horse serum and 100 U penicillin-streptomycin) at 37 °C with 5% CO₂ atmosphere in a humidified incubator. The Flag-tagged PLC- γ stably transfected PC12 cells (Tet-off cell line) were cultured in medium B (85% DMEM, 10% horse serum, 5% FBS, 100 μ g ml⁻¹ G418, 100 μ g ml⁻¹ hygromycin B, 2 μ g ml⁻¹ tetracycline and 100 U penicillin-streptomycin) at 37 °C with 5% CO₂ atmosphere in a humidified incubator. The transfected genes were induced by culturing in medium C (85% DMEM, 10% horse serum, 5% FBS, 100 μ g ml⁻¹ G418, 100 μ g ml⁻¹ hygromycin B and 100 U penicillin-streptomycin) for 24 h. Differentiation was initiated by addition of 50 ng ml⁻¹ NGF with culture medium changed to DMEM with 2% horse serum and 1% FBS, 100 μ g ml⁻¹ G418, 100 μ g ml⁻¹ hygromycin B and 100 U penicillin-streptomycin. Mouse monoclonal anti-haemagglutinin (HA), anti-Myc and protein A and G-conjugated agarose beads were from Calbiochem. We purchased PLC- γ and p110 antibodies from Santa Cruz Biotechnology. Nuclear and cytoplasmic extraction reagents were supplied by Pierce. All chemicals not included above were purchased from Sigma.

Co-immunoprecipitation and *in vitro* binding assays

The experimental procedures for co-immunoprecipitation and *in vitro* binding assays are the same as described in ref. 5. For the GDP-dependent association between Myc-PIKE and GST-PLC- γ -SH223, assays were performed as described²⁵.

GTP loading assays

We performed the assays essentially as described²⁶, with minor modification. PC12 cells were metabolically labelled in phosphate-free Dulbecco's modified Eagle's medium with 0.5 mCi ml⁻¹ [³²P]H₃PO₄ for 3–4 h at 37 °C, and treated with 50 ng ml⁻¹ NGF for the indicated times. The nuclear extract was incubated with 5 μ g rabbit anti-PIKE antibody, and 1:10 volume of PBS/1% BSA/10% charcoal slurry at 4 °C for 30 min, and centrifuged 5 min at 12,000g. Supernatant fluid (0.5 ml) was added to 10 ml protein A Sepharose and rocked for 30 min at 4 °C, then washed 3 times with lysis buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM MgCl₂, 1% Triton X-100) and once with PBS. We performed subsequent steps as described²⁶.

In vitro PI(3)K assay

The nuclear extract was prepared according to the manufacture's protocol. Nuclear PI(3)K was immunoprecipitated with rabbit anti-p110 α from the nuclear extract, and washed with the following buffers: 3 times with buffer A (PBS, 1% NP-40, 1 mM dithiothreitol (DTT)); 2 times with buffer B (PBS, 0.5 M LiCl, 1 mM DTT); and 2 times with buffer C (10 mM Tris-HCl, pH 7.4, 0.1 M NaCl, 1 mM DTT). Subsequent steps were performed as described²⁷.

Nucleotide exchange assays

The assays for dissociation of [³H]GDP from His-PIKE and binding of [³⁵S]GTP γ -S to His-PIKE were performed essentially as described²⁸. The GDP displacement assay was initiated by preloading PIKE with 0.5 μ M [³H]GDP and the dissociation was triggered by adding 0.5 mM unlabelled GTP with 50 nM GST-PLC- γ -SH3 protein or with control buffer. All reactions were carried out at 22 °C in assay buffer (0.1 μ g μ l⁻¹ BSA, 50 nM PIKE,

20 mM Tris, pH 7.6, 1 mM DTT, 5 mM MgCl₂, and 1 mM EDTA) and the aliquots were removed at the indicated times. Dissociation of [³H]GDP was assayed by measuring the decrease in [³H]GDP-PIKE trapped on nitrocellulose filters (HAWP02500 membrane, Millipore). The binding of [³⁵S]GTP γ -S to PIKE was examined in the same assay buffer and at the same temperature as described²⁸. PIKE (50 nM) and 50 nM GST-PLC- γ -SH3 protein were pre-incubated together in assay buffer at 37 °C for 30 min, and the binding initiated by addition of 3 μ M [³⁵S]GTP γ -S (2 μ Ci ml⁻¹). Radioactivity associated with PIKE was measured by filter assay.

Received 18 September; accepted 5 December 2001.

- Bae, S. S. *et al.* Src homology domains of phospholipase C γ 1 inhibit nerve growth factor-induced differentiation of PC12 cells. *J. Neurochem.* **71**, 178–185 (1998).
- Smith, M. R. *et al.* Overexpression of phosphoinositide-specific phospholipase C γ in NIH 3T3 cells promotes transformation and tumorigenicity. *Carcinogenesis* **19**, 177–185 (1998).
- Smith, M. R. *et al.* Phospholipase C- γ 1 can induce DNA synthesis by a mechanism independent of its lipase activity. *Proc. Natl Acad. Sci. USA* **91**, 6554–6558 (1994).
- Huang, P. S. *et al.* An SH3 domain is required for the mitogenic activity of microinjected phospholipase C- γ 1. *FEBS Lett.* **358**, 287–292 (1995).
- Ye, K. *et al.* Pike, A nuclear gtpase that enhances PI3 kinase activity and is regulated by protein 4.1N. *Cell* **103**, 919–930 (2000).
- Pawson, T. & Gish, G. D. SH2 and SH3 domains: from structure to function. *Cell* **71**, 359–362 (1992).
- Gout, I. *et al.* The GTPase dynamin binds to and is activated by a subset of SH3 domains. *Cell* **75**, 25–36 (1993).
- Yu, H. *et al.* Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**, 933–945 (1994).
- Irvin, B. J., Williams, B. L., Nilson, A. E., Maynor, H. O. & Abraham, R. T. Pleiotropic contributions of phospholipase C- γ 1 (PLC- γ 1) to T-cell antigen receptor-mediated signaling: reconstitution studies of a PLC- γ 1-deficient Jurkat T-cell line. *Mol. Cell Biol.* **20**, 9149–9161 (2000).
- Munder, T. & Furst, P. The *Saccharomyces cerevisiae* CDC25 gene product binds specifically to catalytically inactive ras proteins *in vivo*. *Mol. Cell Biol.* **12**, 2091–2099 (1992).
- Lai, C. C., Boguski, M., Broek, D. & Powers, S. Influence of guanine nucleotides on complex formation between Ras and CDC25 proteins. *Mol. Cell Biol.* **13**, 1345–1352 (1993).
- Marmiroli, S. *et al.* Interleukin 1 α stimulates nuclear phospholipase C in human osteosarcoma SaOS-2 cells. *J. Biol. Chem.* **269**, 13–16 (1994).
- Zini, N. *et al.* The intranuclear amount of phospholipase C β 1 decreases following cell differentiation in Friend cells, whereas γ 1 isoform is not affected. *Eur. J. Cell Biol.* **68**, 25–34 (1995).
- Neri, L. M., Borgatti, P., Capitani, S. & Martelli, A. M. Nuclear diacylglycerol produced by phosphoinositide-specific phospholipase C is responsible for nuclear translocation of protein kinase C- α . *J. Biol. Chem.* **273**, 29738–29744 (1998).
- Martelli, A. M. *et al.* Phosphoinositide signaling in nuclei of Friend cells: phospholipase C beta down-regulation is related to cell differentiation. *Cancer Res.* **54**, 2536–2540 (1994).
- Bertagnolo, V. *et al.* Identification of PI-PLC β 1, γ 1, and δ 1 in rat liver: subcellular distribution and relationship to inositol lipid nuclear signalling. *Cell Signal* **7**, 669–678 (1995).
- Bertagnolo, V., Marchisio, M., Volinia, S., Caramelli, E. & Capitani, S. Nuclear association of tyrosine-phosphorylated Vav to phospholipase C- γ 1 and phosphoinositide 3-kinase during granulocytic differentiation of HL-60 cells. *FEBS Lett.* **441**, 480–484 (1998).
- Carpenter, G. & Ji, Q. Phospholipase C- γ as a signal-transducing element. *Exp. Cell Res.* **253**, 15–24 (1999).
- Smith, M. R. *et al.* PLC- γ 1 Src homology domain induces mitogenesis in quiescent NIH 3T3 fibroblasts. *Biochem. Biophys. Res. Commun.* **222**, 186–193 (1996).
- Cherfils, J. & Chardin, P. GEFs: structural basis for their activation of small GTP-binding proteins. *Trends Biochem. Sci.* **24**, 306–311 (1999).
- Quilliam, L. A., Khosravi-Far, R., Huff, S. Y. & Der, C. J. Guanine nucleotide exchange factors: activators of the Ras superfamily of proteins. *Bioessays* **17**, 395–404 (1995).
- Kim, M. J. *et al.* Direct interaction of SOS1 Ras exchange protein with the SH3 domain of phospholipase C- γ 1. *Biochemistry* **39**, 8674–8682 (2000).
- Pei, Z., Maloney, J. A., Yang, L. & Williamson, J. R. A new function for phospholipase C- γ 1: coupling to the adaptor protein GRB2. *Arch. Biochem. Biophys.* **345**, 103–110 (1997).
- Scholler, J. K., Perez-Villar, J. J., O'Day, K. & Kanner, S. B. Engagement of the T lymphocyte antigen receptor regulates association of son-of-sevenless homologues with the SH3 domain of phospholipase C γ . *Eur. J. Immunol.* **30**, 2378–2387 (2000).
- Ye, K., Compton, D. A., Lai, M. M., Walensky, L. D. & Snyder, S. H. Protein 4.1N binding to nuclear mitotic apparatus protein in PC12 cells mediates the antiproliferative actions of nerve growth factor. *J. Neurosci.* **19**, 10747–10756 (1999).
- Rosen, L. B., Ginty, D. D., Weber, M. J. & Greenberg, M. E. Membrane depolarization and calcium influx stimulate MEK and MAP kinase via activation of Ras. *Neuron* **12**, 1207–1221 (1994).
- Ruderman, N. B., Kapeller, R., White, M. F. & Cantley, L. C. Activation of phosphatidylinositol 3-kinase by insulin. *Proc. Natl Acad. Sci. USA* **87**, 1411–1415 (1990).
- Shou, C., Farnsworth, C. L., Neel, B. G. & Feig, L. A. Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing factor for Ras p21. *Nature* **358**, 351–354 (1992).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank G. Carpenter and G. Shawn for some of the plasmids. This work is supported by United States Public Health Service grants (S.H.S.), and Research Scientist Award (S.H.S.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.H.S. e-mail: ssnyder@jhmi.edu.

Structural identification of a bacterial quorum-sensing signal containing boron

Xin Chen*, Stephan Schauder*, Noelle Potier†, Alain Van Dorsselaer†, István Pelczar†, Bonnie L. Bassler* & Frederick M. Hughson*

* Department of Molecular Biology; and † Department of Chemistry, Princeton University, Princeton, New Jersey 08544-1014, USA

† Laboratoire de Spectrométrie de Masse Bio-Organique, Ecole de Chimie, Polymères et Matériaux, 25 Rue Becquerel, 67087 Strasbourg, France

Cell–cell communication in bacteria is accomplished through the exchange of extracellular signalling molecules called autoinducers. This process, termed quorum sensing, allows bacterial populations to coordinate gene expression. Community cooperation probably enhances the effectiveness of processes such as bioluminescence, virulence factor expression, antibiotic production and biofilm development^{1–4}. Unlike other autoinducers, which are specific to a particular species of bacteria, a recently

discovered autoinducer (AI-2)⁵ is produced by a large number of bacterial species. AI-2 has been proposed to serve as a ‘universal’ signal for inter-species communication^{1,2,6,7}. The chemical identity of AI-2 has, however, proved elusive. Here we present the crystal structure of an AI-2 sensor protein, LuxP, in a complex with autoinducer. The bound ligand is a furanosyl borate diester that bears no resemblance to previously characterized autoinducers. Our findings suggest that addition of naturally occurring borate to an AI-2 precursor generates active AI-2. Furthermore, they indicate a potential biological role for boron, an element required by a number of organisms but for unknown reasons.

AI-2 was originally identified in the bioluminescent marine bacterium *Vibrio harveyi* as one of two autoinducers that regulate light production in response to cell density^{5,8}. The synthase required for AI-2 production, LuxS, is widely conserved among Gram-negative and Gram-positive bacteria⁷. AI-2 is produced from S-adenosylmethionine in at least three enzymatic steps (Fig. 1a). Consumption of S-adenosylmethionine as a methyl donor produces S-adenosylhomocysteine, which is subsequently hydrolysed by the nucleosidase Pfs to yield adenine and S-ribosylhomocysteine^{9,10}. Subsequently, S-ribosylhomocysteine is converted to 4,5-dihydroxy-2,3-pentanedione (DPD) and homocysteine^{11,12}. This reaction is catalysed by LuxS^{13,14}. DPD probably forms a cyclic molecule,

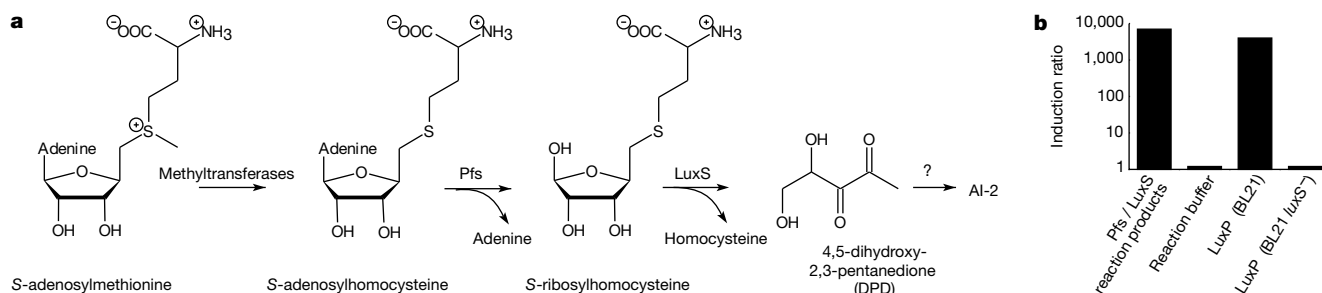


Figure 1 The autoinducer AI-2, synthesized by LuxS, is bound by the sensor protein LuxP. **a**, Biosynthesis of the AI-2 precursor 4,5-dihydroxy-2,3-pentanedione (DPD) from S-adenosylmethionine^{9–13}. **b**, Induction of bioluminescence in the *V. harveyi* bioassay¹³ was measured following the addition of the products of an *in vitro* reaction of

S-adenosylhomocysteine with Pfs and LuxS proteins¹³, reaction buffer, or AI-2 released from LuxP overproduced in LuxS⁺ or LuxS⁻ *E. coli* BL21. Concentrations of AI-2 in the Pfs/LuxS and LuxP (BL21) reactions were estimated to be 20 μ M (see Methods).

Table 1 Phasing and refinement statistics

	Native	KPtCl ₄	K ₂ UO ₂ F ₆	Sm(NO ₃) ₃	Eu(NO ₃) ₃
Resolution (Å)	1.5	2.2	2.5	2.4	2.8
Unique reflections	61,620	17,815	13,657	13,483	9,824
Redundancy	5.5	6	3.2	3.6	3.2
R _{sym} (%)	3.5 (25)	6.3 (32)	9.4 (33)	7.8 (25)	8.5 (29)
Complete (%)	100	100	99.7	99.6	93.1
MIR phasing at 2.8 Å					
Heavy atom sites		3	3	3	2
Phasing power acentric/centric		0.69/0.66	1.35/1.13	0.48/0.42	1.2/0.81
R _{cryst} acentric/centric		0.87/0.85	0.78/0.76	0.95/0.90	0.80/0.82
Anomalous R _{cryst}		0.91	0.78	0.93	n/a
Combined figure of merit MLPHARE/DM ²⁵		0.60/0.82			
Refinement					
Resolution (Å)	25–1.5				
Reflections (>σ)	53,816				
R _{cryst}	0.214 (0.274)				
R _{free}	0.239 (0.317)				
Root-mean-square (r.m.s) deviation					
Bond length	0.007 Å				
Bond angle	1.3°				
Average B factor					
Protein	24.1				
AI-2	11.7				
Water molecules	34.1				
All atoms	25.3				

Data in parentheses are for the highest-resolution shell. $R_{sym} = \sum_i \sum_h |I_i(h) - \langle I(h) \rangle| / \sum_i \sum_h I_i(h)$, where $I_i(h)$ is the i th measurement of h and $\langle I(h) \rangle$ is the mean of all measurements of $I(h)$ for reflection h . Phasing power is $\langle |F_o| \rangle / E$, where $\langle |F_o| \rangle$ is the r.m.s. structure factor amplitude for anomalous scatterers and E is the estimated lack-of-closure error. $R_{cryst} = \sum_i |F_o| - \sum_i |F_c| / \sum_i |F_o|$. $R_{free} = \sum_i |F_o| - \sum_i |F_c| / \sum_i |F_o|$, calculated with a working set of reflections. $R_{test} = R_{cryst}$ calculated with only the test set of reflections, comprising a randomly selected 8% of the data, not used during refinement.

and possibly undergoes further rearrangements, to yield AI-2 (ref. 13).

The detection of AI-2 by *V. harveyi* requires two proteins, LuxP and LuxQ⁸. LuxP belongs to a large family of periplasmic binding proteins whose members bind diverse ligands¹⁵, while LuxQ is a two-component hybrid sensor kinase embedded in the bacterial inner membrane⁸. Several lines of evidence suggest that LuxP is the primary AI-2 receptor. First, LuxP is homologous to the periplasmic ribose-binding proteins of *Escherichia coli* and *Salmonella typhimurium*⁸. This homology is of particular interest as AI-2 is derived from a ribosyl moiety (Fig. 1a). Second, AI-2 is the ligand for a LuxP homologue (LsrB) in *S. typhimurium*¹⁶. Third, purified recombinant LuxP, overproduced in the LuxS⁺ *E. coli* strain BL21,

releases substantial AI-2 signalling activity upon thermal denaturation (Fig. 1b). This activity is nearly equivalent to that produced in an *in vitro* reaction in which *S*-adenosylhomocysteine is incubated with purified recombinant Pfs and LuxS¹³ (Fig. 1b). LuxP overproduced in a LuxS⁻ BL21 strain releases no AI-2 activity upon denaturation. These results demonstrate that LuxP binds AI-2. Furthermore, binding is tight, because the AI-2 remains associated during chromatographic purification. In quorum sensing, the periplasmic LuxP-AI-2 complex probably interacts with LuxQ to transduce the autoinducer signal.

The X-ray crystal structure of recombinant *V. harveyi* LuxP, overproduced in the LuxS⁺ *E. coli* strain BL21, was determined by multiple isomorphous replacement at (MIR) at 2.8 Å resolution and

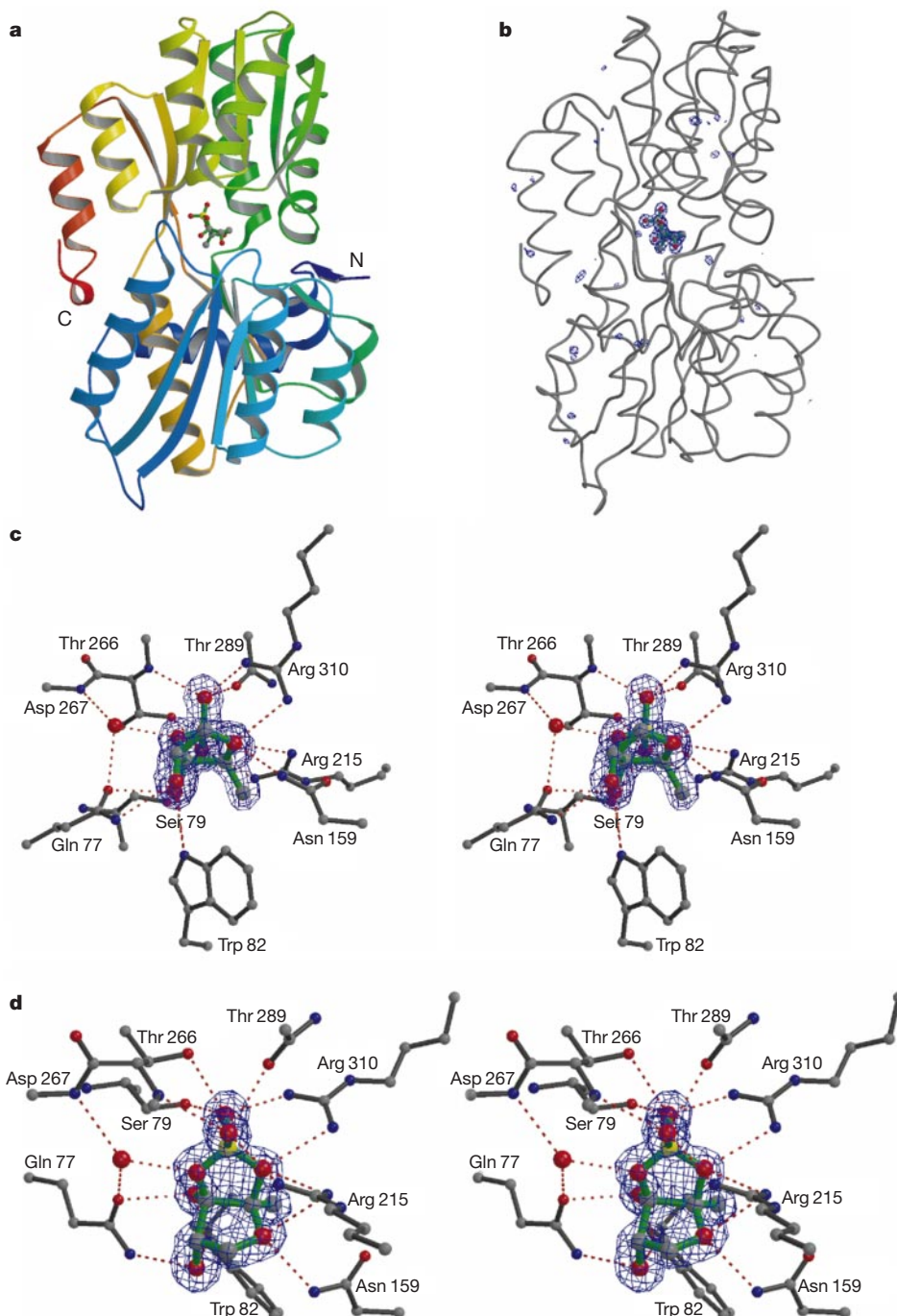


Figure 2 Structure of LuxP-AI-2 complex. **a**, Overview. **b-d**, $F_o - F_c$ difference electron density (contoured at 4σ) calculated using phases derived from the model before AI-2 addition. The final refined model for AI-2 is shown superimposed on this density. Boron,

oxygen, nitrogen and carbon are coloured yellow, red, blue and grey, respectively. In the stereoviews shown in **c-d**, hydrogen bonds are shown as dashed red lines. Figure prepared using O²⁶, Molscript²⁸/Bobsript²⁹ and Raster 3D³⁰.

refined to 1.5 Å resolution (Table 1). Like other periplasmic binding proteins¹⁵, LuxP consists of two similar domains connected by a three-stranded hinge (Fig. 2a). The deep cleft between the two LuxP domains contains additional electron density, corresponding to the AI-2 ligand, that was clearly visible even in initial experimental electron density maps at 2.8 Å resolution. The location of this density is clearly analogous to that of ligands in other periplasmic

binding proteins. The quality of the additional electron density improved steadily as refinement of the protein model progressed until, at 1.5 Å resolution, it was straightforward to construct a model for AI-2 by positioning atoms in the difference electron density (Fig. 2b–d).

Because it is derived from the ribose moiety of *S*-ribosylhomocysteine, AI-2 was initially modelled as a carbohydrate. The assignment of atom type at each position was dictated by both the valence and the local chemical environment within the protein. Thus, tetrahedrally substituted atoms were deemed to be carbons, whereas atoms within hydrogen-bonding distance of at least one (but generally two to three) polar protein atoms or buried water molecules were deemed to be oxygens. Weak bond length and bond angle restraints were applied to the ligand during subsequent refinement. The resulting structural model for the LuxP-AI-2 complex has an R_{cryst} and R_{free} of 0.21 and 0.24, respectively, and exhibits excellent geometry (Table 1). No orientational or chemical heterogeneity of the buried ligand is evident in the crystallographic electron density.

The proposed AI-2 structure contains two fused five-membered rings stabilized within the LuxP binding site by numerous polar interactions (Fig. 3a). Several candidates were considered for the atom bridging the diester. We suggest that it is boron, on the basis of several considerations. Although boron and carbon are not reliably distinguishable on the basis of electron density at this resolution, the presence of carbon at this position would result in a highly unstable orthocarbonate moiety, whereas the borate diester is relatively stable¹⁷. Heavier atoms, such as phosphorus or sulphur, could bridge the diester instead of boron. This possibility can, however, be eliminated because these atoms would give rise to markedly stronger electron density than is observed. In fact, the electron density at this position, although clear, is the weakest of the eight positions in the fused ring structure, consistent with its assignment as boron.

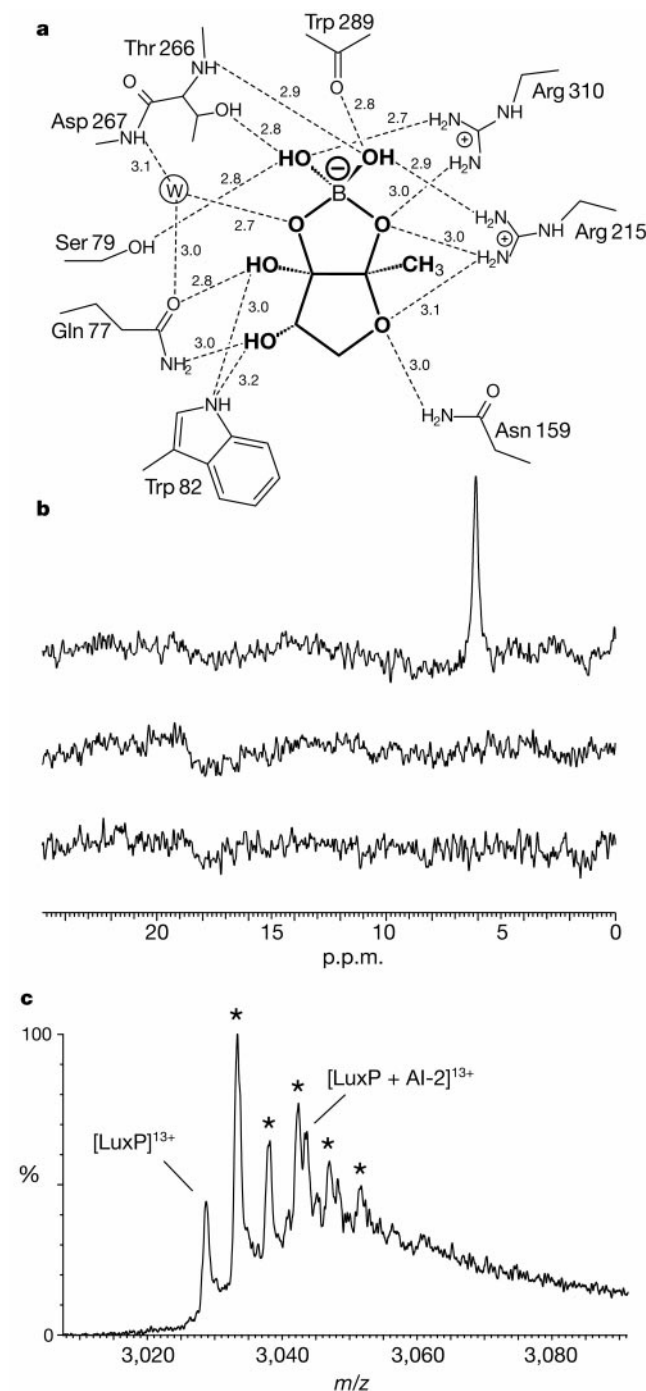


Figure 3 Structure of AI-2. **a**, AI-2 and the hydrogen bond network that stabilizes it in the LuxP binding site. O–O or O–N distances for potential hydrogen bonds are shown in ångströms. **b**, ^{11}B NMR spectra of LuxP overproduced in *E. coli* BL21 (top trace) or BL21 *luxS*[−] (bottom trace). The spectrum of the ultrafiltrate from the BL21 sample is also shown (middle trace). **c**, Mass spectrometric analysis of LuxP-AI-2 in 200 mM ammonium acetate buffer. LuxP alone ($39,295.6 \pm 1.2$), a number of acetate adducts (in increments of 60 daltons; marked with asterisks), and LuxP-AI-2 ($39,489.8 \pm 2.7$) peaks are observed. Approximately 15% of the LuxP retains AI-2. Only acetate adducts are observed in an identical experiment with apo-LuxP (not shown).

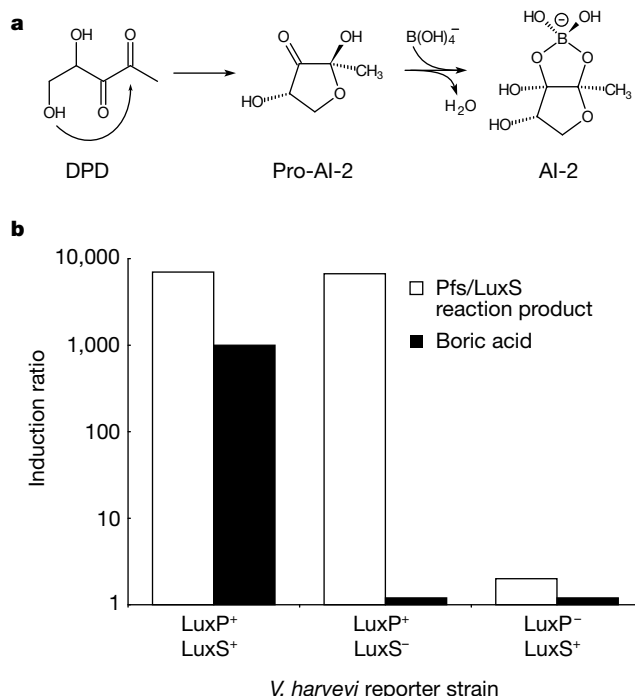


Figure 4 Synthesis of AI-2 from DPD and borate. **a**, Cyclization is thermodynamically favourable. The position of the equilibrium in the second step will depend on the concentrations of pro-AI-2 and borate ($\text{B}(\text{OH})_4^-$). Other possible reactions are not shown. **b**, Light induction by boric acid. The induction of light production in *V. harveyi* reporter strains relative to a buffer control was measured following the addition of *in vitro* Pfs/LuxS reaction products (approximately 20 μM ; white bars) or 1 mM boric acid (black bars). Addition of boric acid does not cause a pH change.

The positively charged side chains of Arg 215 and Arg 310 form hydrogen bonds with three of the four borate oxygens and appear to stabilize the negative charge on the borate (Figs 2c, d and 3a).

To confirm that the ligand contains boron, we carried out ^{11}B NMR spectroscopy. As predicted, the spectrum of the LuxP-AI-2 complex shows a peak signifying the presence of boron (Fig. 3b, top trace). The chemical shift for this peak (6.2 p.p.m.) is within the range (3.9–6.2 p.p.m.) observed for borate esters of carbohydrate 1,2-diols¹⁸. For comparison, borate esters of 1,3-diols and boric acid display chemical shifts of 0.3–0.9 and 18.8 p.p.m. respectively. The ^{11}B NMR peak disappeared after ultrafiltration to remove protein (Fig. 3b, middle trace). Furthermore, it was not observed in apo-LuxP overproduced in LuxS[−] BL21 *E. coli* (Fig. 3b, bottom trace).

To provide additional support for the AI-2 structure proposed here, we determined the mass of the LuxP-AI-2 complex using electrospray ionization mass spectrometry (ESI-MS). Under optimized conditions, peaks are observed for both LuxP and LuxP-AI-2 at molecular masses that differ by 194.2 ± 3.0 daltons (Fig. 3c). This mass increment agrees closely with the molecular mass of the proposed AI-2 molecule (192.9 daltons). We were unable to recover a species corresponding to AI-2 alone.

A chemically straightforward route connects DPD, the product of the LuxS-catalysed reaction, to AI-2 (Fig. 4a). The cyclic form of DPD, which we name pro-AI-2, can react with borate to form a cyclic borate diester. The boric acid required for this reaction is widely available in the biosphere; for example, the average boric acid concentration in sea water is approximately 0.4 mM (ref. 19). (Borate, with a pK_a of 9.2, is present largely as undissociated boric acid at neutral pH.) Furthermore, furanosyl borate diesters similar in structure to AI-2 are relatively stable²⁰. Surprisingly, furanoid rings expected to form stable borate addition products are rare in nature with only two, ribose and apiose, known to have this configuration in their physiological derivatives²¹. In addition to AI-2, other borate adducts may be produced from cyclic derivatives of DPD. Only AI-2, however, is observed in the LuxP crystal structure determined here. Given the large number of specific hydrogen bonds mediating the LuxP-AI-2 interaction (Fig. 3a), it seems unlikely that alternative products could bind with comparable affinity. Taken together, our results provide strong evidence that AI-2 is a cyclic borate diester derivative of DPD. This compound may provide a fruitful lead in the development of antibacterial drugs that target quorum sensing.

Boric acid has a dramatic effect on AI-2 signalling in *V. harveyi*, as shown using a LuxP⁺, LuxS⁺ reporter strain. This strain produces its own AI-2. Endogenously produced AI-2 stimulates light production, but only after a delay during which the AI-2 concentration builds to a threshold stimulatory level. During this delay period, the addition of the activity produced by the *in vitro* reaction of S-adenosylhomocysteine with Pfs and LuxS induces light production. Addition of 1 mM boric acid also results in substantial induction (Fig. 4b). Boric acid concentrations as low as 10 μM caused significant (tenfold) induction (data not shown). Boric acid has no effect on a LuxS[−] strain that cannot synthesize DPD, nor on a LuxP[−] strain that lacks the AI-2 sensor (Fig. 4b). Both of the latter strains show wild-type responses to the other *V. harveyi* autoinducer (AI-1) under these experimental conditions (data not shown). Therefore, the induction of bioluminescence by boric acid is specific for the AI-2 detection system.

We conclude that the availability of borate is limiting under the conditions of the autoinducer bioassay. Additional borate may simply drive the spontaneous production of AI-2 from the available pool of pro-AI-2 (see Fig. 4a). Alternatively, it may provide substrate for an unidentified enzyme that catalyses borate addition. In either case, the strong induction by boric acid in the bioassay is consistent with our proposed chemical structure for AI-2. It will be of particular interest to determine whether borate addition normally occurs immediately after DPD synthesis in the cytoplasm of

the autoinducer-producing cell, extracellularly, or in the periplasm of the recipient cell.

We have shown that AI-2 is a novel furanosyl borate diester. By contrast, previously characterized autoinducers are acyl homoserine lactones, modified oligopeptides or quinolones^{1,2}. One of the most unexpected features of the AI-2 molecule is the presence of a boron atom. Boron has previously been found in a small number of polyketide antibiotics²². It is also known to be essential for vascular plants and several other organisms, including cyanobacteria²¹; however, in no case is its functional role well understood. Our results provide evidence for a biochemically defined function for boron in bacterial quorum sensing.

Whereas acyl homoserine lactone and oligopeptide autoinducers are used for communication within a bacterial species, we have proposed that AI-2 is a universal signalling molecule that facilitates inter-species communication^{1,2}. We have previously shown that LuxS enzymes from a variety of bacteria produce AI-2 activity¹³, presumably through the generation of a common intermediate, pro-AI-2. The results presented here show that the AI-2 sensor in *V. harveyi*, LuxP, binds a borate diester of pro-AI-2. Because the biochemical machinery for synthesizing pro-AI-2 is broadly conserved, and borate is widely available, we speculate that AI-2 is produced and detected by diverse bacterial species. If so, our findings are consistent with the hypothesis that AI-2 is a universal signal.

It remains possible, however, that other pro-AI-2 derivatives are generated and serve as signalling molecules in nature. Alternative derivatives need not necessarily contain boron. Isosteric molecules might also be able to activate signalling via LuxP. Indeed, pro-AI-2 could give rise to diverse signals, their relative abundances dictated by the prevailing environmental conditions and the corresponding availability of reactive species. Integrating inputs from a family of molecules similar in structure to AI-2 could provide bacteria with a rich chemical lexicon with which to communicate.

Note added in proof: A requirement for borate in crosslinking plant cell wall polysaccharides was recently reported³¹. □

Methods

LuxP production

Vibrio harveyi LuxP residues 24–365 (lacking the amino-terminal signal peptide) was overproduced as a glutathione S-transferase (GST) fusion protein. *Escherichia coli* strain BL21, carrying a pGEX4T1-LuxP expression plasmid, was grown in Luria–Bertani medium. Protein expression was induced by the addition of 0.5 mM isopropyl β -D-thiogalactopyranoside. The GST fusion protein was purified by glutathione agarose affinity chromatography. After thrombin cleavage, LuxP contained three additional residues at its N terminus (Gly-Ser-Met). The cleaved product was purified by anion exchange chromatography (MonoQ; Pharmacia). LuxP lacking bound AI-2 was overproduced in a LuxS[−] derivative of BL21 and purified using the same protocol.

Crystallization and diffraction data collection

LuxP (8 mg ml^{−1}) crystallized in 0.1 M Tris-HCl, pH 8.5, 15% PEG 4000, 18% glycerol (v/v) in space group P2₁. Native crystals ($a = 42.3 \text{ \AA}$, $b = 77.5 \text{ \AA}$, $c = 52.0 \text{ \AA}$ and $\beta = 96^\circ$) diffracted to 1.5 $\text{\AA}$ resolution after flash-freezing at 100 K at NSLS beamline X12C. Derivatives (Table 1) were prepared by soaking crystals for 12–18 h in well buffer containing heavy atoms at 10 mM concentration. Derivative diffraction data at 100 K were collected using an R-Axis-IV imaging plate detector mounted on a Rigaku 200HB generator. All data were processed using the HKL program package²³.

Structure determination and refinement

Heavy atom positions were located using SOLVE²⁴ and refined using MLPHARE²⁵. Initial phases calculated with MLPHARE were improved by solvent flattening using DM²⁵ (Table 1). An initial protein model was built into the electron density using the program O²⁶ and refined to 1.5 $\text{\AA}$ resolution using CNS²⁷. No ligand was included in the model until the R_{cryst} and R_{free} had dropped to 0.24 and 0.26, respectively and 187 water molecules (none of them in the ligand position) had been added to the model. No evidence of disorder or multiple conformations was observed for the protein side chains in the AI-2 binding pocket, making it unlikely that the observed ligand electron density represents a mixture of compounds or binding modes. The present model contains residues 27–364 of LuxP, AI-2 and 273 ordered water molecules, and exhibits excellent geometry, with all residues within the most favoured and additional allowed regions of the Ramachandran plot.

Al-2 bioassay

The *in vitro* Pfs/LuxS reaction was carried out on S-adenosylhomocysteine in 10 mM sodium phosphate, pH 7.5, as described¹³. The resulting yield of AI-2 and/or AI-2 precursors was estimated by measuring the amount of homocysteine released¹³. AI-2 activity was assayed using a *V. harveyi* reporter assay⁶. Reporter strains used were BB170 (LuxP⁺, LuxS⁺)⁵, MM30 (LuxP⁺, LuxS⁻)⁷ and BB886 (LuxP⁺, LuxS⁺)⁸. To release AI-2, LuxP was heated for 10 min at 50 °C, then ultrafiltered to remove denatured protein (BioMax Ultrafree, 5,000 relative molecular mass cutoff). The concentration of AI-2 was estimated by assuming complete recovery.

NMR

¹¹B NMR spectra at 4 °C were collected using a Varian Unity/INOVA spectrometer at 160.5 MHz equipped with a 5-mm tunable X/¹H probe (Nalorac), and were referenced indirectly to BF₃O(ET)₂. 80,000 scans were averaged for each spectrum with a 0.25-s recycle time using an approximately 30° flip-angle pulse. LuxP samples contained 0.2 mM protein in 25 mM HEPES, pH 7.5, 100 mM NaCl. Protein was removed from control samples by ultrafiltration (BioMax Ultrafree, 5,000 relative molecular mass cutoff).

Mass spectrometry

Mass spectrometry was performed using an ESI-TOF mass spectrometer (Q-TOF2, Micromass) on LuxP-AI-2 extensively dialysed against 200 mM ammonium acetate, pH 6.5. To help preserve noncovalent complexes, mild interface conditions were used. Specifically, the declustering voltage (V_d), which controls the kinetic energy of the ions in the interface, was set to 20 V. Data acquired in the positive mode were calibrated using a myoglobin standard.

Received 4 September; accepted 14 November 2001.

- Miller, M. B. & Bassler, B. L. Quorum sensing in bacteria. *Annu. Rev. Microbiol.* **55**, 165–199 (2001).
- Schauder, S. & Bassler, B. L. The language of bacteria. *Genes Dev.* **15**, 1468–1480 (2001).
- de Kievit, T. R. & Igilewski, B. H. Bacterial quorum sensing in pathogenic relationships. *Infect. Immun.* **68**, 4839–4849 (2000).
- Kleerebezem, M., Quadri, L. E., Kuipers, O. P. & de Vos, W. M. Quorum sensing by peptide pheromones and two-component signal-transduction systems in Gram-positive bacteria. *Mol. Microbiol.* **24**, 895–904 (1997).
- Bassler, B. L., Wright, M., Showalter, R. E. & Silverman, M. R. Intercellular signalling in *Vibrio harveyi*: sequence and function of genes regulating expression of luminescence. *Mol. Microbiol.* **9**, 773–786 (1993).
- Surette, M. G. & Bassler, B. L. Quorum sensing in *Escherichia coli* and *Salmonella typhimurium*. *Proc. Natl Acad. Sci. USA* **95**, 7046–7050 (1998).
- Surette, M. G., Miller, M. B. & Bassler, B. L. Quorum sensing in *Escherichia coli*, *Salmonella typhimurium*, and *Vibrio harveyi*: a new family of genes responsible for autoinducer production. *Proc. Natl Acad. Sci. USA* **96**, 1639–1644 (1999).
- Bassler, B. L., Wright, M. & Silverman, M. R. Multiple signalling systems controlling expression of luminescence in *Vibrio harveyi*: sequence and function of genes encoding a second sensory pathway. *Mol. Microbiol.* **13**, 273–286 (1994).
- Della Ragione, F., Porcelli, M., Carteni-Farina, M., Zappia, V. & Pegg, A. E. *Escherichia coli* S-adenosylhomocysteine/5'-methylthioadenosine nucleosidase. Purification, substrate specificity and mechanism of action. *Biochem. J.* **232**, 335–341 (1985).
- Cornell, K. A., Swarts, W. E., Barry, R. D. & Riscoe, M. K. Characterization of recombinant *Escherichia coli* 5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase: analysis of enzymatic activity and substrate specificity. *Biochem. Biophys. Res. Commun.* **228**, 724–732 (1996).
- Miller, C. H. & Duerre, J. A. S-ribosylhomocysteine cleavage enzyme from *Escherichia coli*. *J. Biol. Chem.* **243**, 92–97 (1968).
- Duerre, J. A. & Walker, R. D. In *The Biochemistry of Adenosylmethionine* (eds Salvatore, F., Borek, E., Zappia, V., Williams-Ashman, H. G. & Schlenk, F.) 43–57 (Columbia Univ. Press, New York, 1977).
- Schauder, S., Shokat, K., Surette, M. G. & Bassler, B. L. The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum sensing signal molecule. *Mol. Microbiol.* **41**, 463–476 (2001).
- Lewis, H. A. *et al.* A structural genomics approach to the study of quorum sensing: Crystal structures of three LuxS orthologs. *Structure* **9**, 527–537 (2001).
- Quirocho, F. A. & Ledvina, P. S. Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol. Microbiol.* **20**, 17–25 (1996).
- Taga, M. E., Semmelhack, J. L. & Bassler, B. L. The LuxS-dependent autoinducer AI-2 controls the expression of an ABC transporter that function in AI-2 uptake in *Salmonella typhimurium*. *Mol. Microbiol.* **42**, 777–794 (2001).
- Böeseken, J. The use of boric acid for the determination of the configuration of carbohydrates. *Adv. Carbohydr. Chem.* **4**, 189–210 (1949).
- van den Berg, R., Peters, J. P. & van Bekkum, H. The structure and (local) stability constants of borate esters of mono- and di-saccharides as studied by ¹¹B and ¹³C NMR spectroscopy. *Carbohydr. Res.* **253**, 1–12 (1994).
- Bowen, H. J. M. *Trace Elements in Biochemistry* (Academic, London, 1966).
- Mazurek, M. & Perlin, A. S. Borate complexing by five-membered ring vic-diols: vapor pressure equilibrium and N.M.R. spectral observations. *Can. J. Chem.* **41**, 2403–2411 (1963).
- Loomis, W. D. & Durst, R. W. Chemistry and biology of boron. *Biofactors* **3**, 229–239 (1992).
- Schummer, D., Schomburg, D., Irshik, H., Reichenbach, H. & Höfle, G. Absolute configuration and biosynthesis of tartrolon B, a boron-containing macrodiolide from *Sorangium cellulosum*. *Liebigs Ann. Chemie* **1996**, 965–969 (1996).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Terwilliger, T. C. & Berendzen, J. Automated structure solution for MIR and MAD. *Acta Crystallogr. D* **55**, 849–861 (1999).
- Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).

- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brunker, A. T. *et al.* Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 924–950 (1991).
- Esnouf, R. M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132–134 (1997).
- Merritt, E. A. & Bacon, D. J. Raster 3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
- O'Neill, M. A., Eberhard, S., Albersheim, P. & Darvill, A. G. Requirement of borate cross-linking of cell wall rhamnogalacturonan II for *Arabidopsis* growth. *Science* **294**, 846–849 (2001).

Acknowledgements

We acknowledge S. Cooper for suggesting that AI-2 could be a cyclic borate diester. We thank R. Carroll, M. Case, S. Miller, K. Xavier and A. Saxena and the staff of the National Synchrotron Light Source beamline X12C for technical assistance, and J. Carey, C. Eckhart, D. Kahne, C. Lee, Y. Shi, K. Shokat, T. Silhavy, and members of the Hughson and Bassler laboratories for discussions. This research was supported by a Deutsche Akademischer Austauschdienst (DAAD) fellowship (S.S.), NSF and the Office of Naval Research (ONR) (B.L.B.), and NIH (E.M.H.).

Correspondence and requests for materials should be addressed to E.M.H. (e-mail: hughson@princeton.edu). Atomic coordinates for the LuxP-AI-2 complex have been deposited in the Protein Data Bank under the accession code 1JX6.

Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators

Stephen J. Demarest*, Maria Martinez-Yamout*, John Chung*, Hongwu Chen†‡, Wei Xu†, H. Jane Dyson*, Ronald M. Evans† & Peter E. Wright*§

* Department of Molecular Biology and § The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

† Howard Hughes Medical Institute and The Salk Institute for Biological Studies, La Jolla, California 92037, USA

Nuclear hormone receptors are ligand-activated transcription factors that regulate the expression of genes that are essential for development, reproduction and homeostasis¹. The hormone response is mediated through recruitment of p160 receptor coactivators and the general transcriptional coactivator CBP/p300, which function synergistically to activate transcription². These coactivators exhibit intrinsic histone acetyltransferase activity, function in the remodelling of chromatin, and facilitate the recruitment of RNA polymerase II and the basal transcription machinery³. The activities of the p160 coactivators are dependent on CBP. Both coactivators are essential for proper cell-cycle control, differentiation and apoptosis, and are implicated in cancer and other diseases^{4–7}. To elucidate the molecular basis of assembling the multiprotein activation complex, we undertook a structural and thermodynamic analysis of the interaction domains of CBP and the activator for thyroid hormone and retinoid receptors⁸. Here we show that although the isolated domains are intrinsically disordered, they combine with high affinity to form a cooperatively folded helical heterodimer. Our study uncovers a unique mechanism, called 'synergistic folding', through which p160 coactivators recruit CBP/p300 to allow

‡ Present address: Department of Biological Chemistry, Cancer Center, University of California, Davis, California 95817, USA.

transmission of the hormonal signal to the transcriptional machinery.

The complementary interaction domains of mouse CBP (cAMP-responsive (CRE)-binding (CREB) protein) and human ACTR (activator for thyroid hormone and retinoid receptors; also known as RAC3/AIB1/pCIP/TRAM-1) were subcloned, co-expressed⁹ and purified on the basis of previous domain mapping⁸. Co-expression was essential: together the two domains formed a stable stoichiometric complex that could be purified readily, whereas on its own the ACTR domain was highly unstable. Preliminary NMR experiments showed that residues 2,113–2,152 of the original CBP construct, which include several glutamine-proline-glycine repeats, were unstructured in the complex. For NMR structure determination, we therefore used a complex formed from a shortened CBP construct, CBP(2059–2117) containing residues 2,059–2,117, and an ACTR construct, ACTR(1018–1088) containing residues 1,018–1,088. The NMR-derived domain boundaries correspond closely to those determined for the interaction between CBP and the p160 protein SRC1 (ref. 10).

Both the ACTR domain and the CBP domain are intrinsically unfolded in isolation, as indicated by the limited resonance dispersion in ¹H–¹⁵N correlated NMR spectra (Fig. 1a, b). Although circular dichroism spectra show that both domains contain helical secondary structure, with CBP containing more than ACTR, neither unfolds in a cooperative fashion when subjected to temperature or urea denaturation (Fig. 1c). On formation of the complex, however, the two proteins form a single, highly helical and cooperatively folded entity.

The ¹H–¹⁵N NMR spectra of both CBP and ACTR gain significant amide proton dispersion on formation of the complex (Fig. 1a, b). The data for urea denaturation of the ACTR–CBP complex can be fitted to a two-state unfolding model, which shows that the complex unfolds in a single cooperative transition (Fig. 1c). High-affinity binding between CBP and ACTR (dissociation constant (K_d) = 3.4×10^{-8} M, measured by isothermal titration calorimetry; Fig. 1d) is driven by enthalpy ($\Delta H^\circ = -31.7 \pm 1.5$ kcal mol⁻¹). This large enthalpy decrease is partly offset by the high entropic cost ($T\Delta S^\circ = -21.3 \pm 1.5$ kcal mol⁻¹) that is associated with the ACTR and CBP domains folding on complex formation.

We determined the solution structure of the ACTR–CBP complex using restraints derived from multidimensional NMR spectroscopy. ACTR forms an intimate complex with CBP (Fig. 2), burying 1,500 Å² in the molecular interface. The structure comprises six helices, three from the ACTR domain (denoted α 1, α 2 and α 3) and three from CBP (α 1, α 2 and α 3). The overall architecture seems to be new, and DALI searches¹¹ failed to find matches in the database of known structures. Residues 1,018–1,045 and 1,084–1,088 of ACTR are unstructured in the complex. The CBP helices pack in a bundle, exposing a hydrophobic groove between α 1 and α 3 (Fig. 2c). Helix α 1 of ACTR docks in this groove to complete a structural motif that is reminiscent of a four-helix bundle. Helix α 3 in CBP is inclined away from α 2 and extends outward from the helical bundle about two and a half turns. A deep cleft is formed between α 3 and the loop connecting α 1 and α 2 into which helix α 2 of ACTR is inserted (Fig. 2d). Helix α 3 of ACTR packs against an adjacent hydrophobic face of the exposed α 3 helix of CBP. The helices of the ACTR domain are an integral part of the structure, and almost completely encircle α 3 of CBP. Helix α 3 is highly conserved in all CBP/p300 homologues (Fig. 3b) and contributes the largest number of hydrophobic residues to the binding interface. Mutations of residues in the helical regions of the CBP domain, especially in α 3, disrupt binding to SRC1 (refs 10, 12).

Assembly of the two proteins into a complex creates an extensive, leucine-rich hydrophobic core. There are no long-range intramolecular contacts in ACTR itself, and interactions with CBP are required to stabilize its structure. Although the three CBP helices form an intramolecular hydrophobic core, this seems to be too small

and too poorly packed for the protein to form a stable globular fold in the absence of ACTR. We could obtain a stable fold of the isolated CBP domain only at very high salt concentrations (750 mM NaCl or 500 mM Na₂SO₄)—conditions in which it does not bind ACTR. After submission of this manuscript, a structure was reported for a construct similar to our CBP domain, but free in solution¹³. Although helical, the free CBP domain has a different topology to that observed in our complex with ACTR; this structure may represent a single member of the fluctuating ensemble of states that we observe in the free protein.

The bulky hydrophobic residues that constitute the molecular interface occur in clusters in the sequences of the ACTR and CBP binding domains, separated by polar regions that are frequently rich in glycine, serine, proline or glutamine (Fig. 3a, b). These clusters form recognizable motifs with sequence patterns ϕ -X-X- ϕ - ϕ , ϕ - ϕ -X-X- ϕ or a combination of the two, where ϕ is a bulky hydrophobic residue. Many of these hydrophobic motifs are rich in leucine, and it has been suggested that they may function similarly to the L-X-X-L-L motifs that mediate interactions between the p160 coactivators and the ligand-binding domains of nuclear receptors^{12,14}. However, in contrast to the canonical L-X-X-L-L motif, which folds to form an isolated helix on binding to the structured nuclear receptor

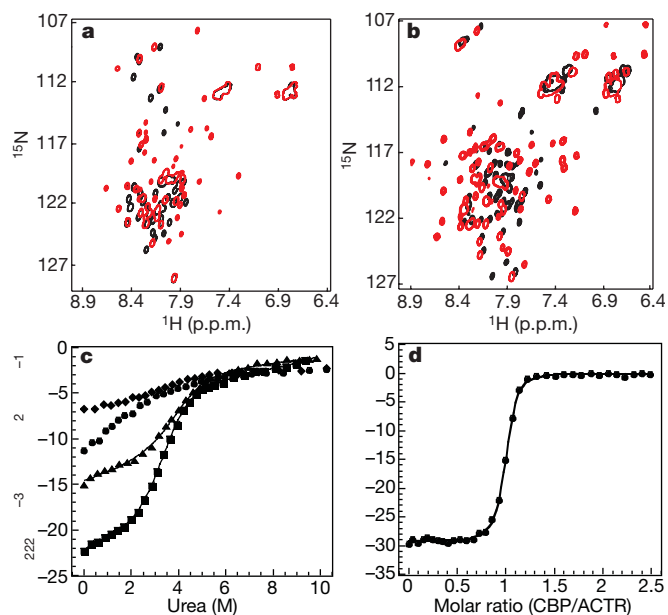


Figure 1 Structural and thermodynamic characterization of free and complexed ACTR and CBP. **a**, ¹⁵N-HSQC (heteronuclear single quantum coherence) spectra of [¹⁵N]ACTR(1018–1088) free (black) and in the presence of excess unlabelled CBP(2059–2117) (red). **b**, [¹⁵N]CBP(2059–2117) free (black) and in the presence of excess unlabelled ACTR(1018–1088) (red). **c**, Urea denaturation of free ACTR(1018–1088) (filled diamonds), free CBP(2059–2117) (filled circles), ACTR(1018–1088)–CBP(2059–2117) (filled triangles) and ACTR(1018–1088)–CBP(2059–2117) (filled squares). Denaturation curves for both complexes were measured in duplicate and each curve was fit separately to a two-state unfolding model³⁰, giving $\Delta G^\circ_U = 3.7 \pm 0.1$ kcal mol⁻¹ and $m = 1,005 \pm 87$ cal mol⁻¹ M⁻¹; where ΔG°_U is the standard free energy change for the unfolding reaction, and M is the dependence of the free energy of unfolding on the concentration of denaturant. Removal of the glutamine-proline-glycine repeat region of CBP (2,118–2,152) did not perturb the stability of the complex; NMR experiments were therefore performed with the CBP(2059–2117) construct. All circular dichroism experiments were carried out with 1–20 μM protein in 2 mM phosphate buffer (pH 6.6), 10 mM NaCl. **d**, Isothermal titration calorimetry (ITC) data for titration of 10 μM ACTR(1018–1088) with 120 μM CBP(2059–2117) in 10 mM Tris buffer (pH 6.9), 50 mM NaCl at 31 °C. The reverse titration gave identical results within experimental error. The stoichiometric constant (n) varied from 0.97 to 1.03 between the two experiments, confirming 1:1 binding.

ligand-binding domain^{15–17}, the disordered leucine-rich motifs in the ACTR and CBP binding domains undergo synergistic folding to create an extensive heterodimerization interface (Fig. 3c).

An additional leucine-rich motif located towards the amino terminus of the ACTR domain, residues 1,029–1,033, does not participate in the hydrophobic core and remains disordered in the complex. This motif is potentially available for interactions with other proteins. Mutation of this region in SRC1 abolishes transcriptional activation dependent on retinoic acid¹², which suggests that it does have a functional role. Notably, the N-terminal leucine-rich motif is not conserved in the transcriptional intermediary factor (TIF) coactivators (Fig. 3a), which may account for possible functional differences compared with SRC1 and ACTR¹⁸. Gene knockout studies have shown only partial compensation of function between different p160 coactivators, suggesting that they have distinct but overlapping roles in receptor activation⁵.

Several polar groups in the molecular interface make complementary interactions that contribute to the specificity of binding. Most notable is a buried salt bridge (Fig. 3d) between Asp 1068 of ACTR ($\alpha 2$) and Arg 2105 of CBP ($\alpha 3$); both residues are conserved completely. Hydrogen-bonding interactions are observed between the side chains of Glu 1045 of ACTR and Asn 2094 of CBP, between Asp 1060 and the γ -oxygen of Ser 2079 (which is a serine or threonine residue in all sequences), and between the backbone

carbonyl of Thr 1059 and the side chain of Lys 2108. The high complementarity between CBP and ACTR ensures tight and specific binding in the face of competition with other transcription factors.

Both ACTR and CBP contain long stretches (up to 26 residues in human ACTR) of glutamine residues or polyglutamine (polyQ) repeats. Expanded polyQ repeats have been implicated in Huntington's disease and other neurodegenerative diseases, and the 15–18-residue polyQ motif in CBP causes it to become sequestered in aggregates of huntingtin¹⁹. It is of note, therefore, that a five-residue polyQ stretch is embedded in the p160-binding domain of CBP (residues 2,082–2,086). This forms a large solvent-exposed loop between $\alpha 1$ and $\alpha 2$ that constitutes one wall of the ACTR-binding groove (Fig. 2c). The glutamine side chains do not seem to make specific interactions with ACTR, however, but project into the solvent where they can potentially interact with other proteins. The polyQ repeat is clearly not crucial for interaction with the p160 proteins because it is not conserved.

The structure of the ACTR–CBP complex provides the first insights into the molecular basis for recruitment of CBP/p300 by the p160 nuclear receptor coactivators. Although coupled folding and binding events have been observed previously in transcriptional activation (reviewed in ref. 20), the interaction between ACTR and CBP represents, to our knowledge, the first example of two incompletely structured domains that fold synergistically to form

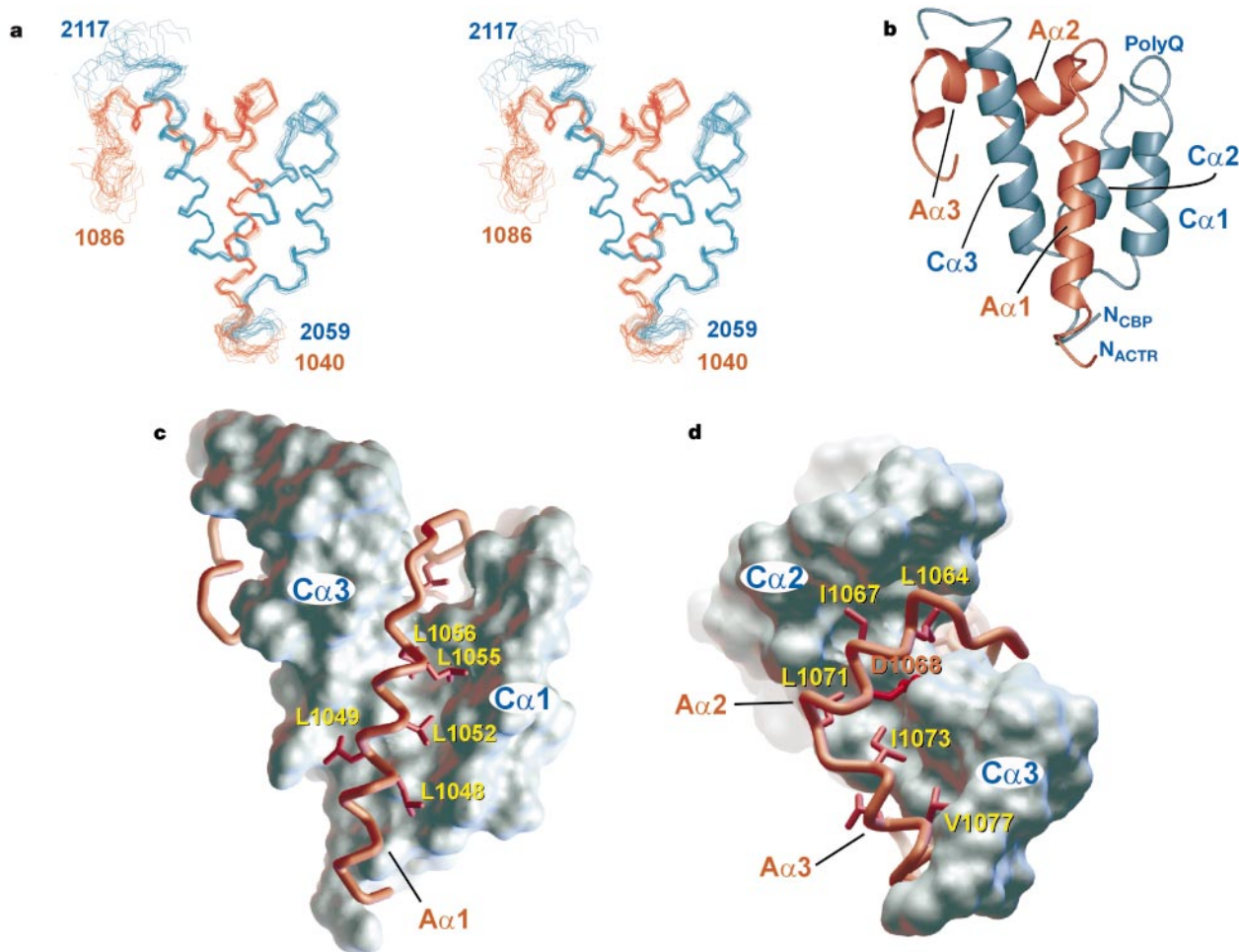


Figure 2 Solution structure of the ACTR–CBP complex. ACTR is pink and CBP blue in all figures. **a**, Stereo view showing best-fit superposition of backbone heavy atoms within the structured region. Residues at the boundaries of the structured region are numbered. **b**, Ribbon representation, in the same orientation as **a**. Helices $\alpha 1$ – $\alpha 3$ and $\alpha 1$ – $\alpha 3$, and the polyglutamine (polyQ) stretch in CBP are labelled. **c**, Surface representation of CBP domain, showing the hydrophobic groove formed by $\alpha 1$ and $\alpha 3$ that accommodates

helix $\alpha 1$ of ACTR. The orientation is the same as in **a** and **b**. Bulky hydrophobic residues from $\alpha 1$ embedded within the groove are labelled. **d**, Surface representation of CBP domain, rotated to show the hydrophobic cleft that binds helix $\alpha 2$ of ACTR. The interactions between $\alpha 3$ and $\alpha 3$ are also shown. Bulky hydrophobic residues of ACTR that form the molecular interface are labelled, as is Asp 1068, which participates in the buried salt bridge.

a tight intermolecular complex. The high affinity of this interaction probably reflects its functional significance for transcription mediated by nuclear hormones and for proper cell-cycle control. The observation that overexpressing ACTR in breast and ovarian cancer cell lines may perturb signal transduction by CBP/p300 (ref. 6) makes the ACTR–CBP complex a potential target for therapy. □

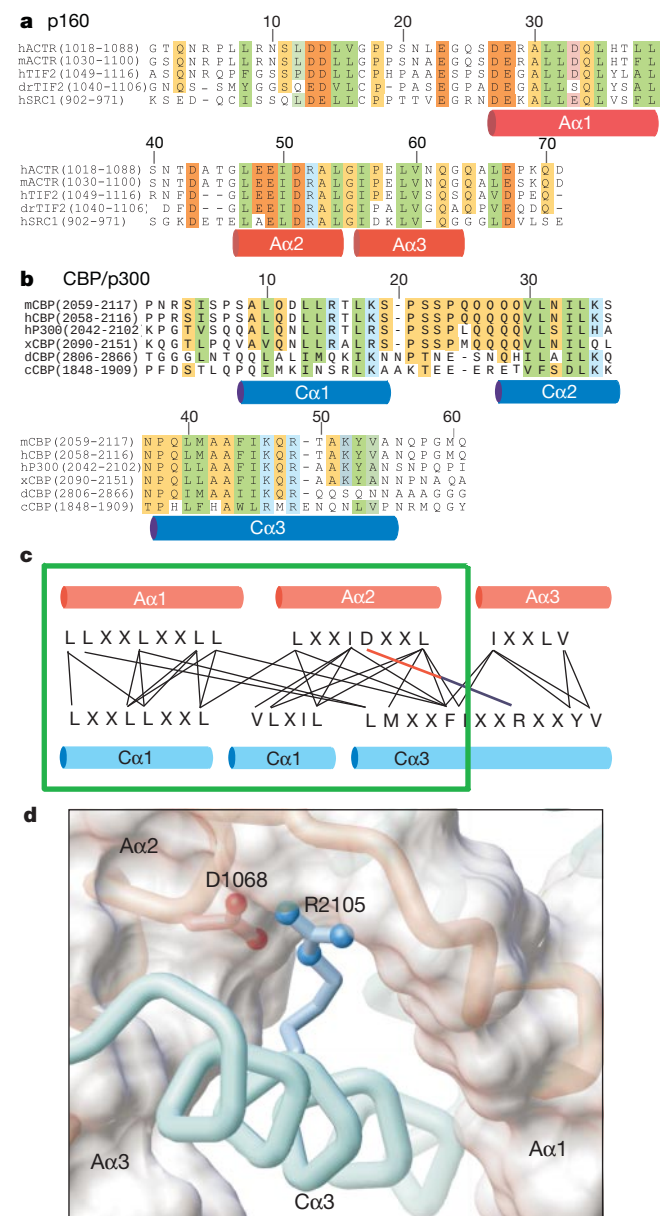


Figure 3 Conserved interactions in the ACTR–CBP complex. **a**, Sequence alignment of the CBP binding domain of human ACTR(1018–1088) and a representative set of p160 coactivators. **b**, Sequence alignment of the ACTR binding domain of murine CBP with other members of the CBP/p300 family. Conserved hydrophobic residues (green), conserved acidic residues (red), conserved basic residues (blue), and other conserved residues (orange) are indicated (h, human; m, murine; x, *Xenopus laevis*; d, *Drosophila*; dr, *Danio rerio*; c, *Caenorhabditis elegans*). **c**, ϕ -X-X- ϕ - ϕ and ϕ -X-X- ϕ hydrophobic contact map defining the interface between ACTR and CBP (ϕ denotes hydrophobic residue). The four ϕ -X-X- ϕ - ϕ motifs that comprise the hydrophobic core are enclosed by a green box. The buried intermolecular salt bridge is indicated. **d**, Close-up of the salt bridge between Arg 2105 and Asp 1068 salt bridge. The solvent-accessible surface of ACTR is shown.

Methods

Preparation of proteins

To avoid protein degradation, the CBP and ACTR interaction domains were co-expressed from a single pET22b-derived vector with two cloning and ribosomal binding sites. The proteins were expressed in Luria broth or minimal medium containing ($^{15}\text{NH}_4$) $_2\text{SO}_4$, $^{15}\text{NH}_4\text{Cl}$ and unlabelled or ^{13}C -labelled glucose. Co-expression yielded $\sim 175 \text{ mg l}^{-1}$ of the protein complex in Luria broth and $\sim 100 \text{ mg l}^{-1}$ in minimal medium. We purified the ACTR and CBP binding domains separately by reverse phase high-performance liquid chromatography, and prepared complexes by stoichiometric recombination of the purified domains. The masses of all proteins were confirmed using matrix-assisted laser desorption/ionization mass spectroscopy. The murine CBP gene was a kind gift from M. Montminy.

NMR spectroscopy

We acquired NMR data on Bruker 500, 600, 750 and 800 MHz spectrometers. All experiments were done at 304 K in 10 mM d_{11} -Tris (pH 6.6), containing 50 mM NaCl. Concentrations varied from 2 to 5 mM for labelled protein samples and from 5 to 20 mM for unlabelled protein. We obtained backbone resonance assignments for complexes made from uniformly labelled [$^{15}\text{N}/^{13}\text{C}$]ACTR(1018–1088) and unlabelled CBP(2059–2152), [$^{15}\text{N}/^{13}\text{C}$]CBP(2059–2152) and unlabelled ACTR(1018–1088), and [$^{15}\text{N}/^{13}\text{C}$]CBP(2059–2117) and unlabelled ACTR(1018–1088), using standard triple-resonance experiments²¹. Torsion angle restraints were derived from two-dimensional 2D CGCO and CGCN, three-dimensional HNHB and HNHA²², and HACAHB-COSY²³ experiments. Distance restraints were obtained using three-dimensional ^{15}N -, ^{13}C - and $^{15}\text{N}/^{13}\text{C}$ -edited nuclear Overhauser enhancement (NOESY) experiments. We carried out 3D ^{13}C -filtered/ ^{13}C -edited NOESY experiments²⁴ to obtain intermolecular distance restraints.

Structure calculations

An initial set of 200 structures was calculated by torsion angle dynamics with the program DYANA²⁵, using 511 unambiguous distance restraints derived from NOEs and 148 torsion angle restraints (65ϕ , 53ψ and $30 \chi_1$). The 100 structures with the lowest target functions were refined by molecular dynamics simulated annealing using the AMBER software package²⁶ and all-atom force field, modified to reduce charges to 20%. Additional NOE distances were added during the AMBER refinement, and partitioned into unambiguous (1,176 total) and ambiguous (707 total) distance restraints using the program SANE²⁷. We chose the best 20 structures to represent the structure of the ACTR–CBP complex. The constraints are well satisfied by the structures, with an average maximum distance violation of $0.23 \pm 0.06 \text{ \AA}$ per structure, and an average torsion angle violation of $4.2 \pm 4.6^\circ$ per structure. We assessed the quality of the structures using PROCHECK²⁸: 90.3% of residues are in the most favoured regions of conformational space, and 8.5% are in additionally allowed regions. Complete structure statistics are provided in the Supplementary Information. Accessible surface areas were calculated with the program MOLMOL²⁹ using a $1.4 \text{ \AA}$ radius probe.

Received 20 August; accepted 28 November 2001.

- Mangelsdorf, D. J. *et al.* The nuclear receptor superfamily: the second decade. *Cell* **83**, 835–839 (1995).
- Chakravarti, D. *et al.* Role of CBP/p300 in nuclear receptor signalling. *Nature* **383**, 99–103 (1996).
- Dilworth, F. J. & Chambon, P. Nuclear receptors coordinate the activities of chromatin remodeling complexes and coactivators to facilitate initiation of transcription. *Oncogene* **20**, 3047–3054 (2001).
- Goodman, R. H. & Smolik, S. CBP/p300 in cell growth, transformation, and development. *Genes Dev.* **14**, 1553–1577 (2000).
- Aranda, A. & Pascual, A. Nuclear hormone receptors and gene expression. *Physiol. Rev.* **81**, 1269–1304 (2001).
- Anzick, S. L. *et al.* AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. *Science* **277**, 965–968 (1997).
- Petrij, E. *et al.* Rubinstein–Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature* **376**, 348–351 (1995).
- Chen, H. *et al.* Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. *Cell* **90**, 569–580 (1997).
- Li, C., Schwabe, J. W., Banayo, E. & Evans, R. M. Coexpression of nuclear receptor partners increases their solubility and biological activities. *Proc. Natl Acad. Sci. USA* **94**, 2278–2283 (1997).
- Sheppard, H. M., Harries, J. C., Hussain, S., Bevan, C. & Heery, D. M. Analysis of the steroid receptor coactivator 1 (SRC1)-CREB binding protein interaction interface and its importance for the function of SRC1. *Mol. Cell Biol.* **21**, 39–50 (2001).
- Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
- McInerney, E. M. *et al.* Determinants of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation. *Genes Dev.* **12**, 3357–3368 (1998).
- Lin, C. H. *et al.* A small domain of CBP/p300 binds diverse proteins. solution structure and functional studies. *Mol. Cell* **8**, 581–590 (2001).
- Voegel, J. J. *et al.* The coactivator TIF2 contains three nuclear receptor-binding motifs and mediates transactivation through CBP binding-dependent and -independent pathways. *EMBO J.* **17**, 507–519 (1998).
- Nolte, R. T. *et al.* Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor- γ . *Nature* **395**, 137–143 (1998).
- Darimont, B. D. *et al.* Structure and specificity of nuclear receptor-coactivator interactions. *Genes Dev.* **12**, 3343–3356 (1998).
- Shiau, A. K. *et al.* The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95**, 927–937 (1998).
- Torchia, J. *et al.* The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function. *Nature* **387**, 677–684 (1997).

19. Nucifora, F. C. *et al.* Interference by huntingtin and atrophin-1 with CBP-mediated transcription leading to cellular toxicity. *Science* **291**, 2423–2428 (2001).
20. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).
21. Clore, G. M. & Gronenborn, A. M. Multidimensional heteronuclear nuclear magnetic resonance of proteins. *Methods Enzymol.* **239**, 349–363 (1994).
22. Bax, A. *et al.* Measurement of homo- and heteronuclear J-couplings from quantitative J correlation. *Methods Enzymol.* **239**, 79–105 (1994).
23. Grzesiek, S., Kuboniwa, H., Hinck, A. P. & Bax, A. Multiple-quantum line narrowing for measurement of H α –H β J coupling in isotopically enriched proteins. *J. Am. Chem. Soc.* **117**, 5312–5315 (1995).
24. Zwaalen, C. *et al.* Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. *J. Am. Chem. Soc.* **119**, 6711–6721 (1997).
25. Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).
26. Pearlman, D. A. *et al.* AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp. Phys. Comm.* **91**, 1–41 (1995).
27. Duggan, B. M., Legge, G. B., Dyson, H. J. & Wright, P. E. SANE (structure assisted NOE evaluation): an automated model-based approach for NOE assignment. *J. Biomol. NMR* **19**, 321–329 (2001).
28. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
29. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51–55 (1996).
30. Myers, J. K., Pace, C. N. & Scholtz, J. M. Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148 (1995).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank M. Allen for technical help and M. Pique for help with computer graphics. This work was supported by grants from the NIH and by the Skaggs Institute for Chemical Biology. S.J.D. was supported by a fellowship from the NIH.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.E.W. (e-mail: wright@scripps.edu). Coordinates have been deposited with the Protein Data Bank under accession number 1KBH.

Structure of the cell-puncturing device of bacteriophage T4

Shuji Kanamaru*[§], Petr G. Leiman*[§], Victor A. Kostyuchenko*[†], Paul R. Chipman*, Vadim V. Mesyanzhinov[‡], Fumio Arisaka[‡] & Michael G. Rossmann*

* Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907-1392, USA

[†] Laboratory of Molecular Bioengineering, Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, 16/10 Miklukho–Maklaya Street, 117997 Moscow, Russia

[‡] Department of Life Science, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8501, Japan
[§] These authors contributed equally to this work

Bacteriophage T4 has a very efficient mechanism for infecting cells¹. The key component of this process is the baseplate, located at the end of the phage tail, which regulates the interaction of the tail fibres and the DNA ejection machine². A complex of gene product (gp) 5 (63K) and gp27 (44K), the central part of the baseplate, is required to penetrate the outer cell membrane of *Escherichia coli* and to disrupt the intermembrane peptidoglycan layer, promoting subsequent entry of phage DNA into the host. We present here a crystal structure of the (gp5–gp27)₃ 321K complex, determined to 2.9 Å resolution and fitted into a cryo-electron microscopy map at 17 Å resolution of the baseplate–tail

tube assembly. The carboxy-terminal domain of gp5 is a triple-stranded β -helix that forms an equilateral triangular prism, which acts as a membrane-puncturing needle. The middle lysozyme domain of gp5, situated on the periphery of the prism, serves to digest the peptidoglycan layer. The amino-terminal, antiparallel β -barrel domain of gp5 is inserted into a cylinder formed by three gp27 monomers, which may serve as a channel for DNA ejection.

Bacteriophage T4 consists of a prolate head with icosahedral ends and a bilayered tail that has a hexagonal baseplate and fibres at its distal end³. The phage recognizes a host by using its long tail fibres, and then anchors the baseplate onto the lipopolysaccharide receptors on the cell surface. Attachment to the cell induces a conformational change in the baseplate that initializes a contraction of the external tail sheath, resulting in penetration of the outer cell membrane and intermembrane peptidoglycan by the rigid internal tail tube². This sequence of events brings the tube into contact with the cytoplasmic cell membrane¹. Subsequently, the phage DNA is injected into the host cell through the tail tube.

The centre of the baseplate, or hub, consists of at least two phage proteins, gp5 and gp27 (refs 4, 5), which form a stable complex *in vivo* and *in vitro*. Phage mutants lacking either of these proteins produce hubless baseplates and, as a result, are tailless and non-infectious. The hub has lysozyme activity attributed to the lysozyme domain of gp5, which is required to digest the cellular intermembrane peptidoglycan layer during phage tail contraction⁶. The gp5 lysozyme domain has 43% sequence identity⁷ with the cytoplasmic T4 lysozyme⁸ (T4L). Unlike T4L, gp5 lysozyme activity is essential for phage growth and infection⁹. When incorporated into the phage baseplate or stored at high concentration, gp5 undergoes maturational cleavage between Ser 351 and Ala 352, but both resultant

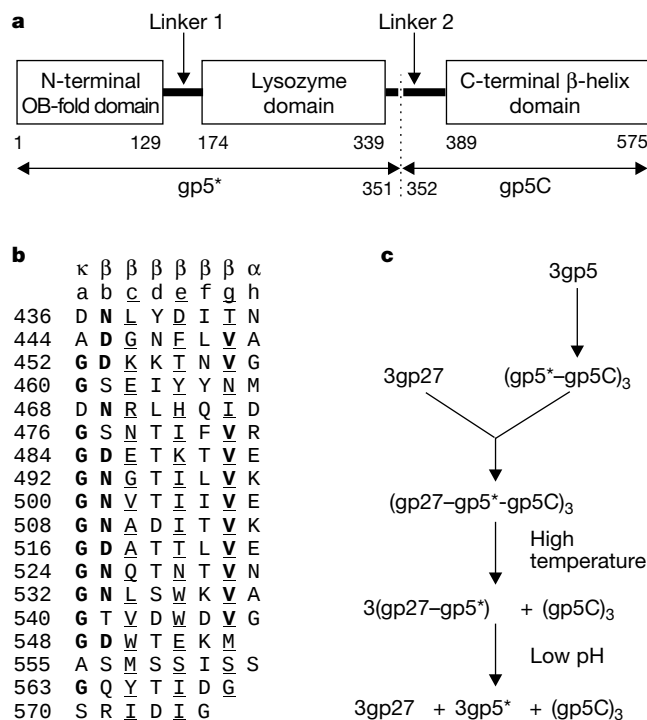


Figure 1 Assembly of (gp27–gp5*–gp5C)₃. **a**, Domain organization within the gp5 monomer. The maturation cleavage is indicated by the dotted line. Initial and final residue numbers are shown for each domain. **b**, Alignment of the octapeptide units composing the intertwined part of the C-terminal β -helix domain of gp5. Conserved residues are in bold; residues facing the inside are underlined. The main-chain dihedral angle configuration of each residue in the octapeptide is indicated at the top by κ (kink), β (sheet), and α (helix). **c**, Assembly of gp5 and gp27 into the hub and needle of the baseplate.

parts, the N-terminal part (gp5*) and the C-terminal part (gp5C) (Fig. 1a), remain in the phage particle¹⁰. gp5C is an SDS-resistant trimer, rich in β -structure¹⁰ and containing 11 VXGXXXX sequence repeats^{7,10} (Fig. 1b). The complex of gp5 with gp27 obtained *in vivo* and *in vitro* contains both gp5* and gp5C, thus comprising a heterononameric assembly: (gp27-gp5*-gp5C)₃ (Fig. 1c). At elevated temperatures, this complex dissociates into three (gp27-gp5*) heterodimers and a (gp5C)₃ homotrimer, showing that gp5C is necessary for trimerization of the entire complex¹⁰ (Fig. 1c). However, dissociation of gp5* from gp27 can occur only at low

pH (S.K., Y. Takeda and F.A., unpublished data). The lysozyme activity of the trimeric gp5*, when stabilized by gp5C, is an order of magnitude less than the activity of the monomeric gp5* (ref. 10).

The structure of the (gp27-gp5*-gp5C)₃ complex (Fig. 2a) resembles a torch (or flashlight) with a length of 190 Å in which the gp27 trimer forms its cylindrical 'head' and the gp5C trimer forms the 'handle'. The gp27 trimer makes a hollow cylinder about 60 Å long, with internal and external diameters of about 30 and 80 Å, respectively, encompassing the three N-terminal domains of gp5* to which the trimeric gp5C 'torch handle' is attached. The

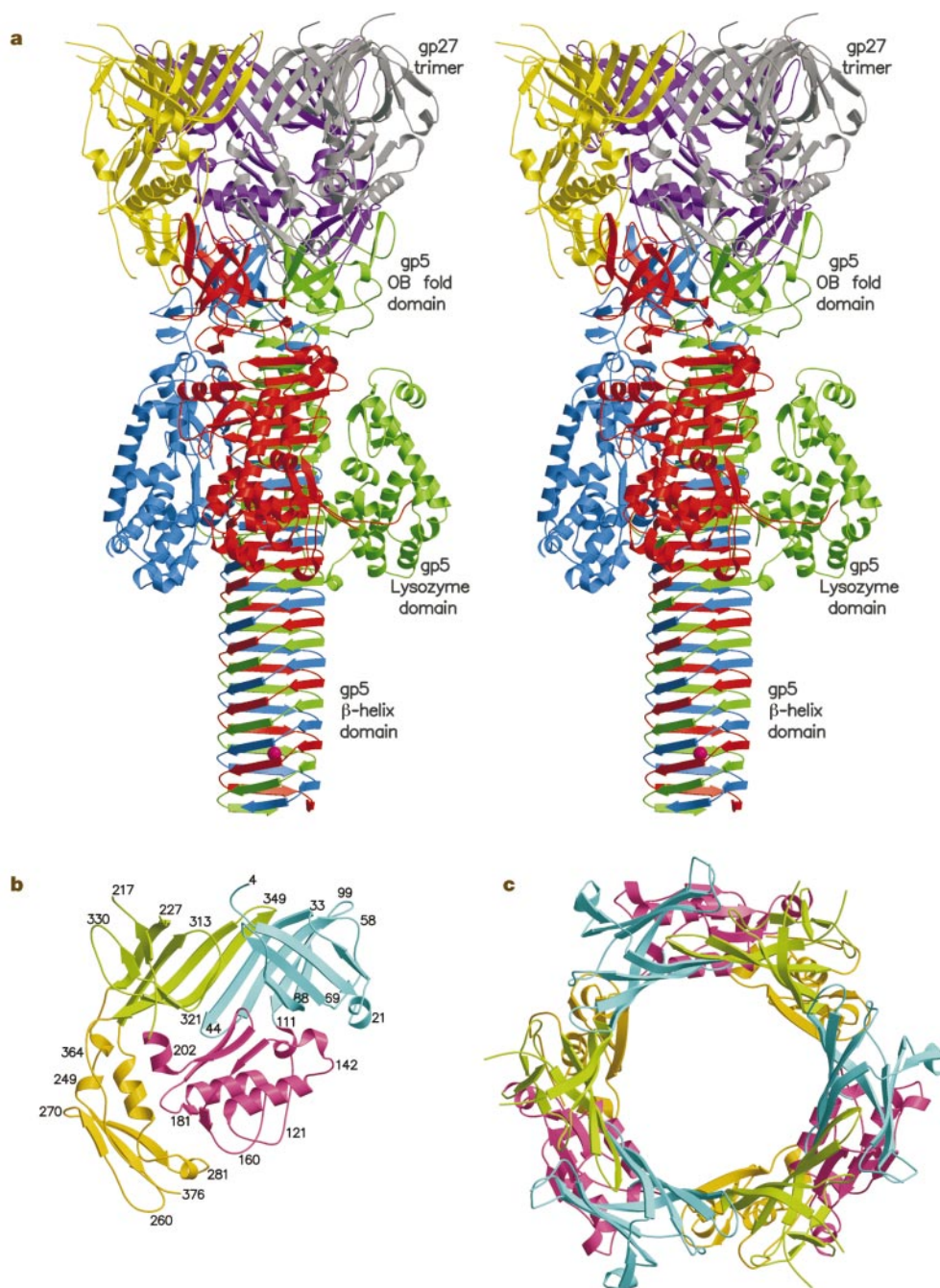


Figure 2 Structure of the (gp27-gp5*-gp5C)₃ complex. **a**, Ribbon stereo diagram of the complex. There is one polypeptide chain of gp5 and one of gp27 in the asymmetric unit. The three-fold symmetry axis of the complex coincided with the crystallographic three-fold axis. The three gp5 monomers are coloured red, green and blue. The three gp27 monomers are coloured yellow, grey and purple. The K⁺ ion within gp5C is shown in pink.

The PO₄ ion is hidden behind the lysozyme domain. **b**, The structure of the gp27 monomer with its four domains, coloured cyan, pink, light green and gold, along the polypeptide chain. **c**, Top view of the gp27 cylinder, showing that the cyan and green domains form a hexagonal torus. Ribbon diagrams were drawn with MOLSCRIPT²² and Raster3D²³.

C-terminal parts of the three gp5C polypeptide chains fold into a β -helical prism in which the faces have a slight left-handed twist. The middle lysozyme domains surround the N end of the prism. Generally, there was good agreement on fitting the (gp27-gp5*-gp5C)₃ structure into a 17 Å resolution three-dimensional image reconstruction of the baseplate-tail tube assembly, obtained with cryo-electron microscopy (cryoEM). As expected, the complex forms the central hub of the baseplate, with the β -helical prism domain forming a needle pointing towards the potential host surface (Fig. 3). The three-fold symmetry of the (gp27-gp5*-gp5C)₃ complex contrasts with the assumed six-fold symmetry of the baseplate, and may, in part, account for the absence of density representing the lysozyme domain. However, it is also probable that the lysozyme is somewhat disordered in its position, made possible by the long, flexible linker regions 1 and 2 (Fig. 1a).

The gp27 monomer, which consists of 319 amino acids¹¹, is folded into four domains (Fig. 2b). The first (residues 2–111) and third (residues 207–239 plus 307–368) domains of the three symmetry-

related gp27 monomers create a torus. Each of these two domains is formed by seven- or eight-stranded antiparallel β -barrels. These can be superimposed on each other, with a root-mean-square deviation (r.m.s.d.) of 2.4 Å between equivalenced C α atoms, by means of a 57° rotation approximately about the crystallographic threefold axis, thus creating a cylinder with a pseudo six-fold symmetry. Although there is no significant sequence similarity (4% sequence identity) between these domains, their external surfaces are mostly hydrophobic, providing a nucleus for the assembly of the six-fold-symmetric baseplate (Fig. 2c). Thus, the gp27 trimer forms an interface between the hexameric baseplate and trimeric gp5. The internal and external surfaces of the gp27 cylinder match the dimensions of the tail tube, suggesting that the gp27 trimer serves as an extension to the 90 Å-diameter tail tube (Fig. 3b).

The second domain (residues 112–206) and the fourth domain (residues 240–306 plus 359–376) of gp27 bind to the two adjacent N-terminal domains of gp5*, thus promoting assembly of the

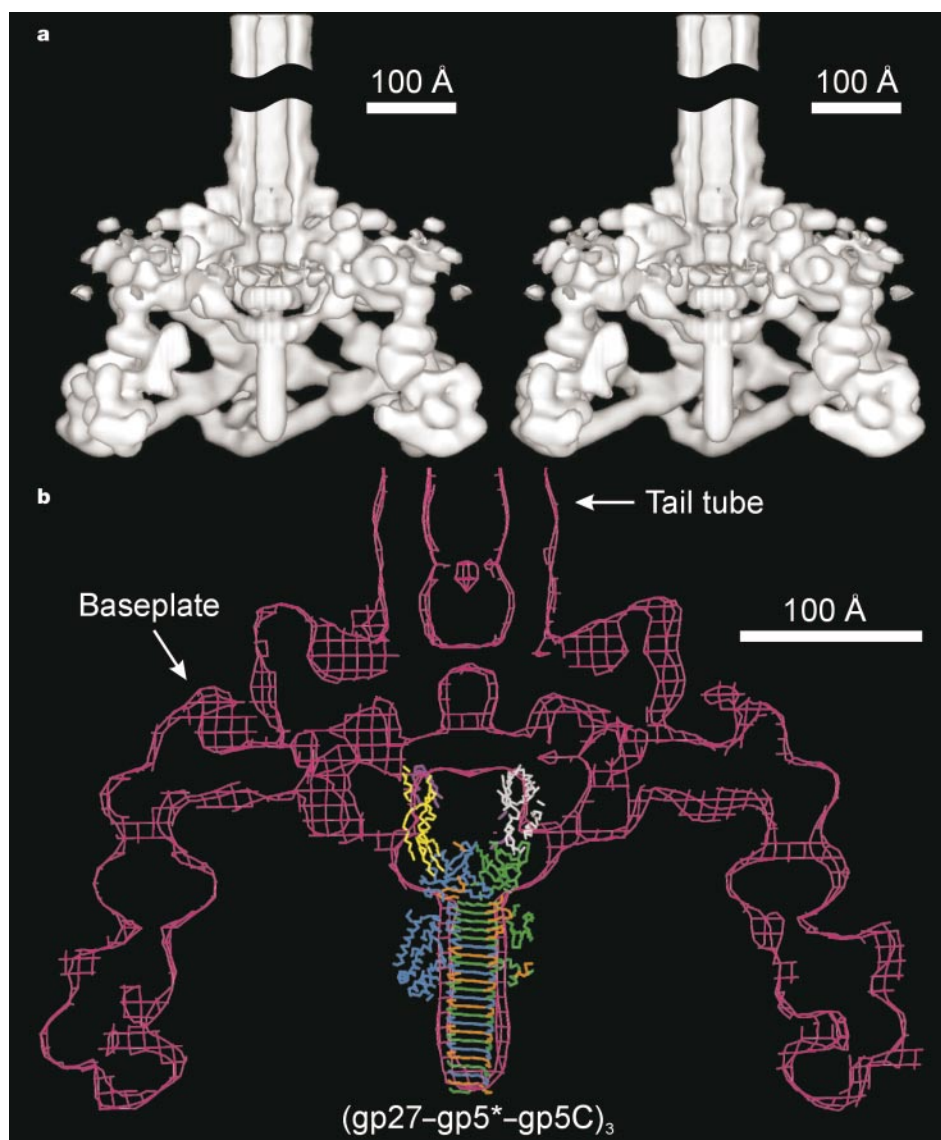


Figure 3 CryoEM reconstruction of the T4 baseplate-tail tube assembly. **a**, Stereo view of the surface-shaded representation. The top quarter of the baseplate has been removed to show the internal features. **b**, Cross-sectioned density fitted with the atomic (gp27-gp5*-gp5C)₃ complex structure. The three-fold-symmetric crystallographic structure fitted into the six-fold averaged cryoEM density is shown in the same colours as Fig. 2a.

Three-fold averaging indicated that the lysozyme domain of gp5 may have moved into the baseplate or may be disordered. For clarity, the contour level of the surface-shaded figure in **a** is higher than the contour level of the cross-sectioned density in **b**. Thus, some of the density representing gp27 in **a** is missing compared with **b**.

oligomeric (gp27–gp5*–gp5C)₃ complex. The interface between gp27 and gp5* is formed mostly by complementary polar and charged residues, with the surface of gp27 being mostly positively charged and the surface of gp5* being mostly negatively charged. The surface contacts between gp27 monomers are formed primarily by a few hydrogen bonds, consistent with the lack of trimer formation when gp27 is expressed by itself (S.K., Y. Takeda and F.A., unpublished data).

Gp5* consists of both an N-terminal and a lysozyme domain (Fig. 1a). The N-terminal domain is a five-stranded, antiparallel β -barrel that has the oligonucleotide/oligosaccharide-binding (OB) fold¹². The OB domain of enterotoxin LT (Protein Data Bank accession number 1LTT) can be superimposed onto the N-terminal domain of gp5 with an r.m.s.d. of 2.4 Å between the 50 structurally equivalenced C α atoms. Unlike the equivalent nucleotide-binding site, the gp5 oligosaccharide-binding site has appropriate aromatic and charged residues, but in the assembled structure of (gp27–gp5*–gp5C)₃, the putative polysaccharide-binding site is occluded by intersubunit contacts. Therefore, the function of the OB domain might be to facilitate binding of gp5* to oligosaccharides of the periplasmic peptidoglycan layer for subsequent digestion by the lysozyme domain, suggesting that gp5* separates from the baseplate during tail contraction (see below).

The gp5 lysozyme domain, located on the periphery of the central β -helix, is connected to the N-terminal domain by linker 1 and to the C-terminal β -helical domain by linker 2, which contains the cleavage site between gp5* and gp5C (Fig. 1a). Seven residues upstream and nine residues downstream of the cleavage site are disordered in the crystal structure. Structural comparison with T4L¹³ shows that residues (Pro 363–Ala 364–Asp 365) of linker 2 of the three-fold-related neighbouring subunit bind, correctly oriented, into three of the four peptide-binding sites used by the peptidoglycan substrate (Fig. 4), but leave the catalytic polysaccharide binding cleft open to solution. The other side of the polysaccharide binding cleft is sterically blocked by the β -helix (Fig. 4). Both of these blockages explain the lack of activity of trimeric (gp5*–gp5C)₃ compared with monomeric gp5* (ref. 10).

The C α atoms of the gp5 lysozyme domain and of T4L can be superimposed on each other with an r.m.s.d. of 1.1 Å, except for five inserted residues in gp5. This analysis also shows that both lysozymes have the same residues in their active sites: Glu 11, Asp 20 and Thr 26 in T4L correspond to Glu 184, Asp 193 and Thr 199 in gp5, respectively, establishing that the enzymatic mechanism is the same

and that the molecules probably have a common evolutionary origin.

The gp5 and T4L enzymes have the same natural substrate, the *E. coli* periplasmic cell wall, the major component of which—(NAG–NAM)–L-Ala–D-iso-Glu–DAP–D-Ala (ref. 13)—contains sugar and peptide moieties. However, gp5 and T4L have activity optima at pH 5.8 and pH 6.8, respectively. The peptide-binding surfaces of gp5 and T4L have several amino-acid differences that cause the peptide-binding surface in T4L to be more positively charged than that in gp5 at neutral pH, possibly accounting for the differences in pH activity optima.

Three gp5C polypeptide chains wind around a central crystallographic three-fold axis to create an equilateral triangular prism that is 110 Å long and 28 Å in diameter. Each face has a slight left-handed twist (–3° per β -strand), as is normally observed in β -sheets. The width of the prism face narrows gradually from 33 Å at the N end to 25 Å at the C end, thus creating a pointed needle. This narrowing is caused by a decrease in size of the external side chains and by the internal methionines 554 and 557, which break the octapeptide repeat near the tip of the helix (Fig. 1c). The first five β -strands (residues 389–435) form an antiparallel β -sheet, creating one of the three faces of the prism. A similar β -sheet prism domain has been observed in the central portion of the phage P22 tailspike protein^{14,15}. However, unlike the hydrophobic interior of the P22 tailspike, the gp5 structure is mostly polar in its interior. The succeeding 18 β -strands of gp5 form a three-start β -helix together with the other 2 three-fold-related polypeptides, generating six complete turns of the β -cylinder. The intertwined, C-terminal part of the β -helical prism (residues 436–575) is a remarkably smooth continuation of its three monomeric N-terminal parts (residues 389–435). Another example of a three-fold intertwined fibrous β -structure is the adenovirus fibre¹⁶. However, the lack of superimposable structural similarity suggests that these viral fibres may be the consequence of convergent evolution to useful, stable, oligomeric fibrous structures.

The octapeptide sequence of the helical intertwined part of the prism (residues a–h, Fig. 1b) has dominant glycines at position a, asparagines or aspartic acids at position b, valines at position g, and polar or charged residues at position h (Fig. 1b). Residues b–g form extended β -strands (Ramachandran angles $\phi \approx -129^\circ$, $\psi \approx 128^\circ$) that run at an angle of 75° with respect to the helix axis. The glycines at position a ($\phi \approx -85^\circ$, $\psi \approx -143^\circ$, an allowed region of the Ramachandran diagram) and residues at position h ($\phi \approx -70^\circ$, ψ

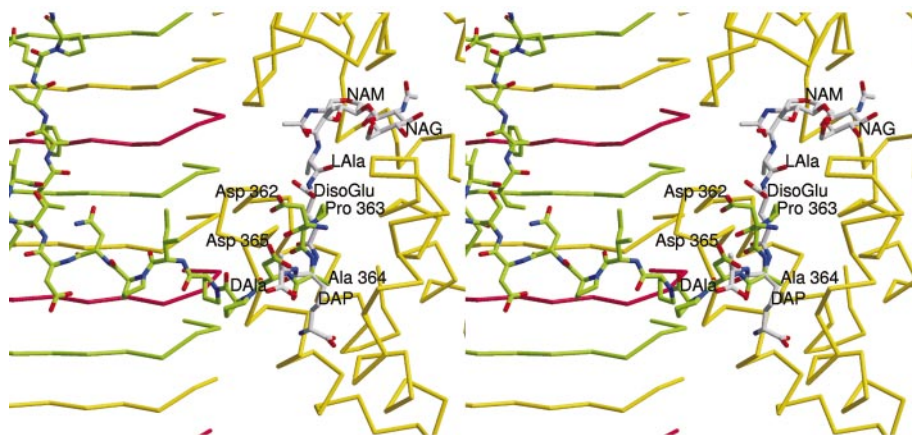


Figure 4 Structural comparison with T4L¹³ shows that residues Pro 363–Ala 364–Asp 365 of linker 2 of the three-fold-related neighbouring subunit bind, correctly oriented, into three of the four peptide-binding sites used by the peptidoglycan substrate. Polypeptide chains of gp5 are represented by virtual C α –C α bonds, except for those in linker 2, which are represented as a stick model, with green for carbon atoms. The

peptidoglycan substrate as bound to T4L¹³ is also shown as a stick model, with grey for carbon atoms. The three individual polypeptide chains of gp5 are coloured yellow, red and green. DAla, D-Ala; LAla, L-Ala; DisoGlu, D-iso-Glu; NAG, N-acetylglucosamine; NAM, N-acetylmuramic acid; DAP, diaminopimelic acid.

$\approx -30^\circ$, typical for α -helices) kink the polypeptide chain by about 130° clockwise. The conserved valines at position g always point to the inside of the β -helix and form a 'knob into holes' arrangement with the main-chain atoms of the glycines at position a and the aliphatic part of the side chains of residues at position c. Asp 436 replaces the normal glycine in position a and is at the start of the β -helix. This substitution may be required for folding of the β -helix, because the Asp 436 O_δ atom makes a hydrogen bond with O_γ of Ser 427 from a three-fold-related polypeptide chain. The side-chain oxygen atoms of Asp 468, which also occupies position a, form hydrogen bonds with residues in the lysozyme domain.

The inside of the β -helix is progressively more hydrophobic toward its C-terminal tip (Fig. 1b). The middle part of the helix has a pore, which is filled with at least 42 water molecules bound to polar and charged side chains. The helix is stabilized by two ions situated on the helix symmetry axis: a phosphate ion coordinated by three Lys 454 residues and a potassium ion coordinated by three Glu 552 residues. These ions were identified by the presence of density peaks and their amino-acid environments. Most probably, these ions are incorporated into the β -helix during folding, as no such ions are present in the purification and crystallization solutions. All these features of the β -helix create a highly stable structure, consistent with its resistance to 10% SDS or 2 M guanidine HCl. The surface of the β -helix is highly negatively charged. This charge is necessary to repel the phosphates of the lipid bilayer when the β -helix penetrates through the outer cell membrane.

The (gp27-gp5*-gp5C)₃ complex can be fitted into the cryoEM image reconstruction of the T4 baseplate-tail tube assembly (Fig. 3b). The gp27 cylinder is shown to be an extension of the tail tube, which is continued by the three N-terminal domains of gp5. The latter associates with the N end of the β -helix that terminates the extended tube. The three lysozyme domains are located around the β -helix under the baseplate, but vary in their exact positions. The diameter of the gp27 cylinder is of a size that can accommodate a double-stranded DNA helix, thus forming a channel that would allow the DNA to penetrate the tail tube as far as the N-terminal domain of gp5.

On attachment of the baseplate to the cell surface, the tail sheath contracts, exerting a force onto the tail tube towards the cell membrane. This force would then be transmitted through the gp27 cylinder and the N-terminal domain of gp5 to the β -helix, causing the latter to puncture the outer membrane. As the contraction of the tail sheath progresses, the β -helix would be able to span the entire $40 \text{ \AA}$ width of the outer membrane, thereby enlarging the pore in the membrane. Subsequently, the three lysozyme domains would reach the peptidoglycan layer to digest the cell wall, allowing penetration of the tail tube to the inner membrane for injection of the phage DNA into the host. □

Methods

Separate expression vectors containing gene 5 and gene 27 and a coexpression vector were constructed (S.K., Y. Takeda and F.A., unpublished data). The complex was purified with the aid of a gp5 C-terminal His tag. Crystals were formed in hanging drops at 12°C from 5.2% PEG 8000, 0.65 M Tris buffer at pH 8.5, and 35% glycerol. They belonged to space group R32, with cell dimensions of $a = 138.1$ and $c = 388.9 \text{ \AA}$. Selenomethionine (SeMet)-substituted gp5 was insoluble, but, fortunately, the SeMet-substituted gp27 was soluble. A chimaeric complex containing separately expressed native gp5 and the SeMet-substituted gp27 was purified and crystallized. Four-wavelength multiple anomalous dispersion data were collected using the chimaeric crystals of the complex that diffracted to $2.9 \text{ \AA}$ resolution. Fifteen of the sixteen selenium atoms in the gp27 monomer were found with the help of the program SHELX¹⁷. These sites were refined and used to obtain phases with the program MLPHARE¹⁸. The phases were then improved by solvent flipping using the program CNS¹⁹. The resultant electron density map was easily interpreted with the graphics program XtalView²⁰. Refinement of the atomic model with the program CNS¹⁹ converged to the working R -factor of 21% and R_{free} of 28%.

Mutant T4 phages lacking the ability to assemble the tail sheath and head were used to produce the baseplate-tail tube assembly. Three-dimensional image reconstructions of the baseplate-tail tube assembly were computed by SPIDER²¹. The baseplate was assumed to be six-fold symmetric. A total of 418 images were employed to obtain a map at about $17 \text{ \AA}$ resolution.

Received 1 October; accepted 3 December 2001.

- Goldberg, E., Grinius, L. & Letellier, L. in *Molecular Biology of Bacteriophage T4* (ed. Karam, J. D.) 347–356 (American Society for Microbiology, Washington, DC, 1994).
- Coombs, D. H. & Arisaka, F. in *Molecular Biology of Bacteriophage T4* (ed. Karam, J. D.) 259–281 (American Society for Microbiology, Washington, DC, 1994).
- Eiserling, F. A. & Black, L. W. in *Molecular Biology of Bacteriophage T4* (ed. Karam, J. D.) 209–212 (American Society for Microbiology, Washington, DC, 1994).
- Kikuchi, Y. & King, J. Genetic control of bacteriophage T4 baseplate morphogenesis. III. Formation of the central plug and overall assembly pathway. *J. Mol. Biol.* **99**, 695–716 (1975).
- Vanderslice, R. W. & Yegian, C. D. The identification of late bacteriophage T4 proteins on sodium dodecyl sulfate polyacrylamide gels. *Virology* **60**, 265–275 (1974).
- Nakagawa, H., Arisaka, F. & Ishii, S. Isolation and characterization of the bacteriophage T4 tail-associated lysozyme. *J. Virol.* **54**, 460–466 (1985).
- Mosig, G., Lin, G. W., Franklin, J. & Fan, W. H. Functional relationships and structural determinants of two bacteriophage T4 lysozymes: a soluble (gene e) and a baseplate-associated (gene 5) protein. *New Biol.* **1**, 171–179 (1989).
- Matthews, B. W. & Remington, S. J. The three dimensional structure of the lysozyme from bacteriophage T4. *Proc. Natl Acad. Sci. USA* **71**, 4178–4182 (1974).
- Takeda, S., Hoshida, K. & Arisaka, F. Mapping of functional sites on the primary structure of the tail lysozyme of bacteriophage T4 by mutational analysis. *Biochim. Biophys. Acta* **1384**, 243–252 (1998).
- Kanamaru, S., Gassner, N. C., Ye, N., Takeda, S. & Arisaka, F. The C-terminal fragment of the precursor tail lysozyme of bacteriophage T4 stays as a structural component of the baseplate after cleavage. *J. Bacteriol.* **181**, 2739–2744 (1999).
- Bova, R. et al. Bacteriophage T4 gene 27. *Nucleic Acids Res.* **18**, 3046 (1990).
- Murzin, A. G. & Chothia, C. Protein architecture: new superfamilies. *Curr. Opin. Struct. Biol.* **2**, 895–903 (1992).
- Kuroki, R., Weaver, L. H. & Matthews, B. W. A covalent enzyme-substrate intermediate with saccharide distortion in a mutant T4 lysozyme. *Science* **262**, 2030–2033 (1993).
- Steinbacher, S. et al. Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. *Science* **265**, 383–386 (1994).
- Seckler, R. Folding and function of repetitive structure in the homotrimeric phage P22 tailspike protein. *J. Struct. Biol.* **122**, 216–222 (1998).
- van Raaij, M. J., Mittraki, A., Lavigne, G. & Cusack, S. A triple β -spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401**, 935–938 (1999).
- Sheldrick, G. M. in *Crystallography of Biological Macromolecules* International Tables for Crystallography Vol. F (eds Rossmann, M. G. & Arnold, E.) 734–738 (Kluwer Academic, Dordrecht, 2001).
- Otwinski, Z. in *Isomorphous Replacement and Anomalous Scattering. Proc. CCP4 Study Weekend, 25–26 January 1991* (eds Wolf, W., Evans, P. R. & Leslie, A. G. W.) 80–86 (Science and Engineering Research Council, Daresbury, UK, 1991).
- Brünger, A. T. et al. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- McRee, D. E. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).
- Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
- Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).

Acknowledgements

We thank T. S. Baker for establishing the electron microscopy facilities at Purdue, where we collected the cryoEM data, A. A. Simpson for advice and help in data collection, and B. W. Matthews for discussion of the results. We thank the staff of BioCARS for their help and advice in the data collection at the Advanced Photon Source beam lines 14-BM-C and 14-BM-D. We thank S. Wilder for help in preparation of the manuscript. The work was supported by a National Science Foundation grant to M.G.R.; Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan to F.A.; a Howard Hughes Medical Institute grant to V.V.M.; a reinvestment grant from Purdue University; and a Keck Foundation award for the purchase of a Philips CM300 electron microscope.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.G.R. (e-mail: mgr@indiana.bio.purdue.edu). The coordinates of the (gp27-gp5*-gp5C)₃ complex have been deposited in the Protein Data Bank, under an accession number of 1K28.

Technological trends

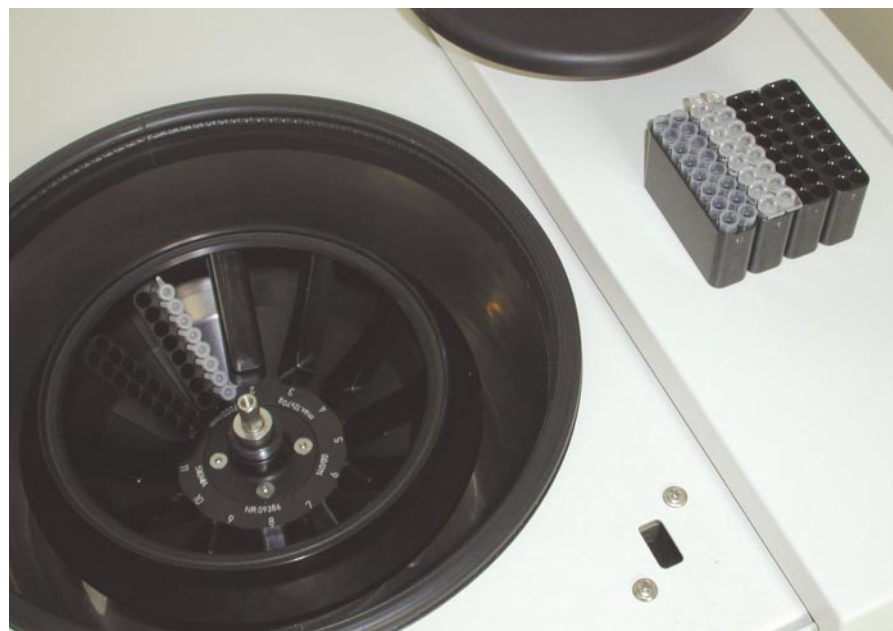
Products with a biotech bias, including a separator good for over 100 l.p.h.

HST-Rotor

Retrotech www.retrotech-online.de

High speed, high throughput

The HST-Rotor system from Retrotech consists of a fixed-angle rotor, adapter cassettes, tubes and sealing accessories. It can be used for sample preparation of small biological specimens such as viruses, cellular organelles, nucleic acids, proteins and macromolecular complexes by centrifugation of up to 192 samples at up to approximately 40,000g. Sample volumes of 300 to 2,000 ml can be transferred rapidly from single tubes, microplates or deep well plates in a microplate eight-strip format to special eight-strip centrifugation tubes loaded into rotor cassettes in a microplate 2 × 8 strip format. The rotor fits Sigma 3K30 and Beckmann Avanti T30 high-speed table-top centrifuges. Depending on application, sample preparation and centrifugation can be completed within 30 min to 2 hours. Applications include transfusion medicine, HIV viral load quantification, virus research and other areas requiring high-speed centrifugation and high sample throughput.



High-throughput centrifugation from Retrotech.

Stericup-VP filters

Millipore www.millipore.com

Remove mycoplasma from tissue culture media

Millipore claims that the Stericup-VP filter unit is the first vacuum filter cup that both removes mycoplasma and provides high-

speed filtration of up to 1 litre with low protein binding. The Stericup-VP filter cup features 0.1 µm Millipore Express membrane, which has been proven to remove mycoplasma. The membrane's asymmetric structure is designed for high filtration capacity and speed. The unit is suitable for use in laminar flow hoods and the "no tip/easy grip" flask is designed to make gripping the receiver easier. As an added convenience, the bottom of the receiver flask is slightly recessed, enabling capped flasks to be stacked for convenient storage. Mycoplasmas are a family of very small bacteria without a rigid cell wall. The cell wall deficiency allows them to pass through a 0.22 µm pore size bacterial retentive membrane filter.



Stericup: catch those mycoplasma.

PicoPure

Arcturus www.arctur.com

RNA isolation kit

The PicoPure kit for recovery of RNA from laser capture microdissection (LCM) cells is optimized to recover the maximum amount of total cellular RNA while maintaining RNA integrity and recovering even low-abundance mRNA. The kit delivers RNA in as little as 10 µl of buffer, ready for use in gene expression studies. Isolated RNA may then be amplified with the company's RiboAmp RNA amplification kit to produce microgram quantities ready for labelling and hybridization onto microarrays. The kits comes with all buffers and purification columns and a three-step isolation protocol. Each kit contains sufficient supplies for 10 isolations.

In situ gene expression analysis

Innogenex www.innogenex.com

Good in parts

The latest gene expression catalogue from Innogenex features tissues and tissue microarrays, GeneXPRO contract services, proprietary immunodetection systems and a broad selection of antibodies and probes. Tissue microarray is a collection of histologically characterized multiple tissues from different human and animal samples on a charged slide. A variety of kits are offered for the detection of antibodies and probes in cells and tissues. Complete kits for *in situ* hybridization and immunohistochemistry on rodent or non-rodent tissues contain pre-titrated, ready-to-use reagents. Ancillary reagents and detection system components are also available separately.

Instant ELISA kits

MedSystems Diagnostics GmbH

www.instant-elisa.com

Undiluted pleasure

Microtitre plates containing freeze-dried, pre-mixed quantities of the standard protein, coating antibody, streptavidin HRP and sample diluent forms the basis of Medsystems' instant ELISA kits. Standard protein titration as well as reagent dilution and addition are eliminated; the number of washings required is also reduced. The assay procedure is straightforward: add water and sample, incubate and wash, add TMB, and stop the reac-

tion. Instant ELISAs are available for several TNF receptors, interleukins, CD markers, ICAMs and Selectins.

GeneLinker Gold 2.0

Molecular Mining Corporation

www.molecularmining.com

Out this week

Released on 21 January, MMC's GeneLinker Gold version 2.0 Gene Expression Analysis Software adds several new features in addition to speed and data handling improvements. Enhancements include 3-dimensional Principal Components Analysis plots, summary statistics and profile matching. Features carried over from previous versions include Self Organizing Maps and Jarvis-Patrick clustering. GeneLinker Gold is recommended for users who need a gene expression analysis tool that is easy to use, comprehensive, and affordable. Rather than an annual licensing fee, GeneLinker Gold is licensed in perpetuity to the end user. Software upgrades and tech support are free of charge for a full year. After that users can choose to use the version they have or purchase upgrades and support at a discount.

DNMT1

Affinity Bioreagents

www.bioreagents.com

Antibody to DNA methyltransferase 1

Part of a collection of DNA methylation antibodies that includes antibodies to Dnmt3A, Dnmt3b and DNA methyltransferase 1 associated proteins, DNMT1 has been optimized for Western blot applications in human and mouse species. DNMT1 DNA methyltransferases methylate the 5-position of cytosine in the context of CpG dinucleotides. DNA methylation plays a key role in development, imprinting and tumorigenesis.

AnaLight Bio250

Farfield Sensors

www.farfield-sensors.com

Make light work of biomolecular analysis

Optical interrogation of biomolecular interactions without recourse to labelling is the aim of Farfield's AnaLight Bio250 technology. The

system utilizes polarized light from a laser passing down a waveguide onto which a biological layer has been deposited. The evanescent field emanating out of the waveguide interrogates the biological layer. By examining the resulting optical interference caused by molecular interactions in the biological layer, real-time analysis can be performed. Applications include the study of protein interactions and metabolism and high-speed pharmacokinetic measurements.

Virtual laboratory

ThermoHybaid

www.hybaidd.com

Oligonucleotides online

Modified oligonucleotides can be ordered online from ThermoHybaid's virtual lab. RP-HPLC purification is included. Long and complex DNA and RNA oligos and chimaeras can be provided, and the company offers advice on synthesis strategies via its website. Mass spectroscopy analysis is completed for each batch of ISO9001 certified oligonucleotides, which are suitable for use in real-time PCR, SNP analysis, antisense research and spotting on microarrays.

PRISM SnaPshot

Applied Biosystems

www.appliedbiosystems.com

Multiplex system for SNP analysis

As many as 23,000 single nucleotide polymorphism (SNP) genotypes can be analysed daily with the SnaPshot system for five-dye SNP analysis. Up to 10 primer/template combinations can be multiplexed simultaneously in a single tube. The system is suitable for use with all ABI PRISM 310, 3100 and 3700 instruments. Genotypes of the multiplexed samples can be assigned using Genotyper v.3.7 software with GeneScan-120 LIZ Size Standard.

A.C.E. Phosphate Buffers

AMRESCO

www.amresco-inc.com

Phosphate buffers optimized for capillary electrophoresis

A.C.E. Phosphate Buffers are recommended for the separation of various biomolecules, including proteins. The buffers are pre-mixed and ready-to-use in formulations optimized for capillary electrophoresis use. A.C.E. Phosphate Buffers are manufactured in nuclease-free water, and all buffers are 0.2 µm filtered during packaging.

LAPX 404

Alfa Laval

www.alfalaval.com

Made to scale

The LAPX 404 separator is designed to allow biochemists to create laboratory-scale biotech processes and then accurately scale them up to production levels. The unit incorporates controls, piping and a centrifugal



The wheel thing: Alfa Laval's LAPX 404.

separator into a skid with a footprint of less than one square metre. It can be wheeled into place and connected to utilities, making it suitable for development programmes involving different laboratories or teams of technicians. A low-shear disk inlet ensures that product treatment is gentle and the risk of anti-foam fouling is eliminated. The LAPX 404 can be used with a wide variety of biological substances since the risk of blinding and subsequent flux reduction is avoided. An intermittent self-cleaning feature increases solids handling capacity. Nominal maximum capacity is 1,000 l.p.h, but the company expects that in a task such as cell harvesting of bacterial fermentation broth, typical throughput will be 100–200 l.p.h., with a handling limit of 33 l.p.h for solids.

EmulsiFlex-C1000

Avestin

www.avestin.com

Good mixer

The EmulsiFlex-C1000 is equipped with a pneumatically controlled homogenizing valve and can be used to process product at a rate of 500 to 1000 l.p.h at operating pressures up to 30,000 p.s.i. The product path, which does not contain any O-rings or gaskets, is steam-in-place (SIP) sterilizable. All wetted parts are autoclavable and can be disassembled for inspection. The homogenizer is suitable for pharmaceutical and biotechnological applications such as production of emulsions, nanoparticles and liposomes as well as cell rupture.

These notes are compiled in the Nature office from information provided by the manufacturers.



Speed merchant: AnaLight Bio250.

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:

Leonie Welss (4954)
European Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy:

Nevin Bayoumi (4978)

Scandinavia:

Sille Opstrup (4994)

Spain/ Portugal:

Leonie Welss (4954)

Production Manager:

Billie Franklin
To send films and materials use
London address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner
Tel + 49 89 54 90 57 11/2
Fax + 49 89 54 90 57 20
e-mails: p.phelan@nature.com
k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@nature.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakatomachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@naturejp.com
Japan Manager: Kate Cowan

naturejobs

Battling the brain drain

For European nations, retaining scientific talent is difficult. And getting it back once the leading lights have left the country is even harder. But a scheme by Britain's Royal Society is proving that neither goal is impossible. In less than a year, the Royal Society-Wolfson Research Merit Award has secured places for 20 researchers at British universities.

In a recent round, three of the awards attracted scientists working in the United States back to the United Kingdom. These include a British expert in quantum chemistry who left England when he was 13 (see Movers, back page).

Although the award's administrators don't explicitly state that their aim is to entice British scientists back, they quietly celebrate such victories. "It's a bonus if there's a candidate from abroad," says Charlotte Ennis, manager of research appointments for the Royal Society.

The award is attractive because it provides up to a further £75,000 (US\$107,000) per year over five years — up to half of which can be applied as a salary supplement and the remainder as research expenses. After five years, the institution can determine ways to continue its support.

Universities apply for the award by nominating someone they already have on staff or whom they wish to recruit (see www.royalsoc.ac.uk/funding/index.html). There are now four rounds of awards each year, with up to seven appointments per round.

If the scheme becomes a permanent fixture, it could make a sizable impact on British science. And perhaps European countries that suffer from even more serious brain drains (see *Nature* **415**, 245; 2002) will adopt similar schemes.

Paul Smaglik

Naturejobs editor



Contents

MOVERS

Neuroscientist heads for industry; new director for space biomedicine centre; quantum chemist comes home; and more [Back page](#)

WWW.NATUREJOBS.COM

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

Transitions

X Paul-Pierre Pastoret will succeed **Chris Bostock** as director of Britain's Institute for Animal Health later this year. Pastoret is currently head of the immunology and vaccinology unit in the faculty of veterinary medicine at the University of Liège, Belgium.

X Stef Heylen will join Dutch antibody and vaccine company Crucell as vice-president of development and regulatory affairs on 1 March. Before joining Crucell, Heylen was vice-president of medical affairs for Europe, Middle East and Africa at Janssen-Cilag. He has worked for the Janssen Research Foundation since 1987 in various senior management positions.

X Last November Genome Prairie, an initiative on wheat and canola genomics backed by the Canadian government, appointed **Randal Johnston** as its first president and chief executive. Johnston gave up the directorship of the Southern Alberta Cancer Research Centre and a post as associate vice-president of research at the University of Calgary to take up the new job.

X Linguist Jean-Marie Hombert this month succeeded **Marie-Claude Maurel** as director of the department of humanities and social sciences at the CNRS, the French scientific research agency. Hombert has been a professor of linguistics at Lumière-Lyon 2 University since 1980.

X Christoph Leemann succeeds **Hermann Grunder** as director of the Thomas Jefferson National Accelerator Facility in Newport News, Virginia. Leemann has been interim director since Grunder left to head the Argonne National Laboratory in 2000.

SPACE MEDICINE

Jeffrey Sutton didn't have a clear destination in mind as he navigated a career path that switched from medical training to graduate work in neuroscience to a PhD in theoretical physics. "I never really knew where it would lead — except I was having a lot of fun along the way," he says.

But his varied career helped to prepare him for his new position as director of the National Space Biomedical Research Institute in Houston, Texas, a consortium aimed at developing solutions to medical problems encountered during space flights. His background will help him to examine areas such as the effects of deep-space radiation on a spacecraft's occupants and the impact of zero gravity on bone and muscle.

A key aim of the consortium is to design small, non-invasive ways to diagnose and treat medical problems. Sutton already has experience of medical-technology development with his previous joint appointments as head of the neural-systems group at the Massachusetts General Hospital and associate professor in the Harvard-MIT Division of Health Sciences and Technology. Including the effects of space adds another exciting variable. That challenge attracted him to the consortium in 1999 and is now drawing proposals from a number of high-profile scientists, including Nobel laureates.

"People want to be players," Sutton says. "Some of it is impact. Some of it is legacy. And some of it is pure fun." Although Sutton officially joined the consortium as director in November, he will move to Houston in June.

CHEMISTRY

After "a very long absence", British quantum chemist **Peter Taylor** is returning to England, thanks to a scheme designed to counteract the UK brain drain.

Taylor, currently deputy director of the San Diego Supercomputer Center at the University of California, San Diego, accepted a professorship in the University of Warwick's chemistry department. He will help to build up the university's recently established Centre for Scientific Computing — a smaller, more focused version of the San Diego centre.

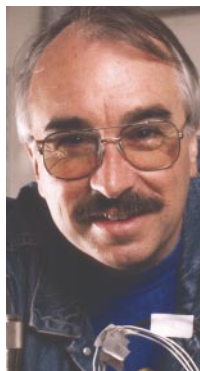
Taylor left England at the age of 13, when his parents moved to Australia. He has spent the past 10 years at San Diego.



Jeffrey Sutton



Dennis Choi



Peter Taylor

Although he'll miss the south California weather, he's keen to return to England, which he already visits regularly. A Royal Society-Wolfson Research Merit Award, which will supplement his pay, helped to attract him back.

Biologist **Timothy Karr** of the University of Chicago and **Gabriel Aeppli** of the NEC Research Institute in Princeton, New Jersey, also received the award. Karr is joining the University of Bath's department of biology and biochemistry. Details of Aeppli's new post have not yet been released.

PHARMACEUTICALS

Dennis Choi this month joins another academic-turned-industrial-scientist, **Peter Kim** (see Movers, *Naturejobs* 12 July 2001), at Merck Research Laboratories in West Point, Pennsylvania.

Choi is leaving the Washington University School of Medicine in St Louis, Missouri, where he served for over 10 years as head of the neurology department and as Andrew B. and Gretchen P. Jones professor. He was also neurologist-in-chief at Barnes-Jewish Hospital in St Louis, Missouri, and founding director of the Center for the Study of Nervous System Injury at the Washington School of Medicine.

He is now executive vice-president, neurosciences, at Merck Research Laboratories and reports to Kim, who joined the company as executive vice-president, research and development, last year.

Choi, whose research interest is in understanding the cellular mechanisms that underlie brain or spinal-cord injury in neurological disease states, made the move so that he could think more broadly about diseases of the nervous system. He has also always wanted to make an impact with therapies and he feels that he has a better opportunity in industry.

"I see this move as a logical extension of my career-long interest in translational research," says Choi. "I've always been focused on the bridge." He notes that the bridge between academia and the drug-development industry is more tempting to cross now because drug development is becoming increasingly based on defined mechanisms rather than on large-scale screening of compounds against possible targets. As a result, academic and industrial career paths are becoming more "intertwined", says Choi.

CONTACTING US AT MOVERS

Please send any information on appointments or job moves to:
 ◆ naturejobseditor@naturedc.com